AD-767 403

THE PRACTICAL IMPACT OF THE RECENT COMPUTER ADVANCES ON THE ANALYSIS AND DESIGN OF LARGE SCALE NETWORKS

Howard Frank

Network Analysis Corporation

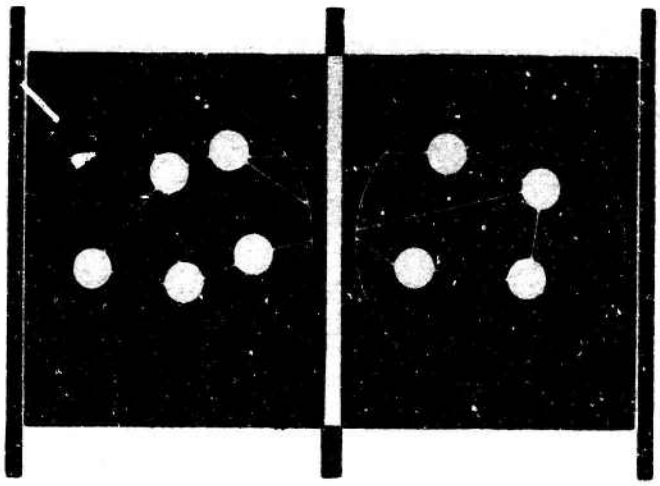
Prepared for:

Advanced Research Projects Agency

May 1973

The Practical Impact of Recent Computer Advances on the Analysis and Design of Large Scale Networks

First Semiannual Technical Report



nac

NETWORK ANALYSIS CORPORATION

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Network Analysis Corporation
Beechwood, Old Tappan Road
Glen Cove, New York 11542

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
None

3. REPORT TITLE
First Semiannual Technical Report, May 1973, for the Project "The Practical Impact of the Recent Computer Advances on the Analysis and Design of Large Scale Networks"

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
First Semiannual Report, May 1973

5. AUTHOR(S) (Last name, first name, initial)
Network Analysis Corporation

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1973	~~191~~ 198	11

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
DAHC15-73-C-0135	Semiannual Report 1B
b. PROJECT NO.	
c. ARPA Order No. 2286	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	---

10. AVAILABILITY/LIMITATION NOTICES
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Advanced Research Projects Agency, Department of Defense

13. ABSTRACT New research results on the following major questions are reported: Growth and cost performance tradeoffs for the ARPANET, traffic sensitivity analyses with measured traffic and with possible ILLIAC IV traffic effects and ARPANET capacity expansion. Initial cost, delay, throughput and reliability analyses for a 1000 node packet switched network based on ARPANET technology. Development of efficient algorithms for location and line layout for local and regional access to Terminal Interface Processors. The development of an interactive data handling system based on an IMLAC display and distributed ARPANET computation. Study of flow control, routing and time delay and system organization for broadcast packet systems.

14. KEY WORDS
Computer networks, throughput, cost, reliability, survivability, ARPA Computer Network, store-and-forward, packet switching.

Security Classification

First Semiannual Technical Report

May 1973

For the Project

The Practical Impact of Recent Computer Advances on the Analysis and Design of Large Scale Networks

Principal Investigator
and Project Manager:

Howard Frank (516) 671-9580

ARPA Order No. 2286

Contractor: Network Analysis Corporation

Contract No. DAHC15-73-C-0135

Effective Date: 13 October 1972

Expiration Date: 12 October 1973

Sponsored by

Advanced Research Projects Agency
Department of Defense

SUMMARY

Technical Problem

Network Analysis Corporation's contract with the Advanced Research Projects Agency has the following objectives:

- To determine the most economical and reliable configurations to meet growth requirements in the ARPANET.
- To study the properties of packet switched computer communication networks.
- To develop techniques for the analysis and design of large scale networks.
- To determine the cost/throughput/reliability characteristics of large packet switched networks for application to Defense Department computer communication requirements.
- To apply recent computer advances, such as interactive display devices and distributed computing, to the analysis and design of large scale networks.

General Methodology

The approach to the solution of these problems has been the simultaneous

- study of fundamental network analysis and design issues.
- development of efficient algorithms for large scale network analysis and design.
- development of an interactive distributed display and computational system to deal with large scale problems.
- application of the new analysis and design tools to study cost and performance tradeoffs for large systems.

Technical Results

In this report, we present new results on the following major questions:

- Growth and cost performance tradeoffs for the ARPANET, traffic sensitivity analyses with measured traffic and with possible ILLIAC IV traffic effects and ARPANET capacity expansion.

- Initial cost, delay, throughput and reliability analyses for a 1000 node packet switched network based on ARPANET technology.

- Development of efficient algorithms for location and line layout for local and regional access to Terminal Interface Processors.

- The development of an interactive data handling system based on an IMLAC display and distributed ARPANET computation.

- Study of flow control, routing and time delay and system organization for broadcast packet systems.

Department of Defense Implications

The Defense Department has vital need for highly reliable and economical communications. The results of this reporting period have established the validity of packet switching for users with the massive data communications problems such as the DOD. The analyses indicate that a major portion of the cost of implementing this technology will occur in providing local access to the networks. Hence the development of effective local and regional communication techniques must be given high priority. Fundamental system considerations, and routing, and flow control techniques for the promising technique of broadcast packet radio are described in the report.

Implications for Further Research

Further research must continue to develop tools for the study of large network problems. These tools must be used to investigate tradeoffs between terminal and computer density, traffic variations, the effects of improved local access schemes such as packet radio, the use of domestic satellites in broadcast mode for backbone networks and the effect of link and computer hardware variations in reliability on overall network performance. The potential of these networks to the DOD establishes a high priority for these studies.

TABLE OF CONTENTS

TABLE OF CONTENTS Continued

Page

1. ARPANET RELIABILITY ANALYSIS AND ENHANCEMENT

1. INTRODUCTION

Node and link failures in a communication network tend to reduce network throughput and to interrupt communications between node pairs. The performance degradation is generally measured using the following criteria:

- Probability of network disconnected (P_{nc})
- Fraction of disconnected node pairs (F_{nc})
- Average network throughput (as opposed to the maximum throughput obtained with perfectly reliable network components)

One of the goals in the design of ARPANET is to provide a satisfactory reliability, in terms of the three above criteria.

In recent study, NAC has evaluated the reliability of the present ARPANET configuration (early 1973), using the actual node and link failure rates, as from the NCC cumulative statistics. This evaluation is summarized in Section 2. The results indicate that the present configuration is not sufficiently reliable. NAC therefore has investigated the possible ways of improving reliability at minimum cost and without drastic modifications to the existing topology.

The first technique experimented by NAC was the introduction of 4.8 Kbs links, to make the topological structure more robust by eliminating eventual pendant nodes and very unreliable long chains.

The results of this experiment, described in Section 3, show that the addition of a few 4.8 Kbs links can improve considerably both P_{nc} and F_{nc}. However, the average throughput remains practically unchanged due to the small capacity of the links. Throughput improvement is obtained at a much higher cost with insertion of 50 Kbs links.

A second technique for reliability improvement consists of the installation of by-pass switches at some of the IMP's and is discussed in this report. Two of the lines incident to an IMP are connected through a by-pass. The by-pass is activated with a switch when the IMP goes down and saves the connection (and the throughput) between two neighbor nodes.

Sections 4, 5 and 6 discuss the effect of by-pass switches on ARPANET reliability and throughput. Section 7 combines the results of low speed line and by-pass switch utilization to provide a recommended network configuration and its reliability analysis.

2. RELIABILITY ANALYSIS OF ARPANET

The reliability of the current (based on December 1972), 33-node ARPANET shown in Figure 1.1* is first evaluated. Node and link failure rates were obtained from the BBN report of January 1973: the link statistics are based on a 2-year period (January 1971 - December 1972), the node statistics on a 6-month period (June 1972 - December 1972). A copy of the BBN statistics appears in the Appendix to this chapter. The reliability analysis was performed with the analysis program described in the NAC Third Semiannual Report for Contract DAHC 15-70-C-0120.

For the current ARPANET (with no additional links), the fraction of disconnected node pairs and the probability of network disconnected are given in Table 1.1. The probability of network disconnected is very high ($\cong$ 0.20), and can be attributed to two main contributions: the pendant node CCA (notice the very high failure rate ($\cong$ 0.10) of the link BBN-CCA) and the long chains in the topological configuration. The first contribution can be easily evaluated by eliminating the pendant node. This reduces the probability of disconnection to 0.12 but has only a minor effect on the expected fraction of node pairs not communicating. As for the second contribution, notice that the longer the chains,

*Notice that, in this study, the link ABER-BELV was removed, since it has been affected in the past by serious technical troubles.

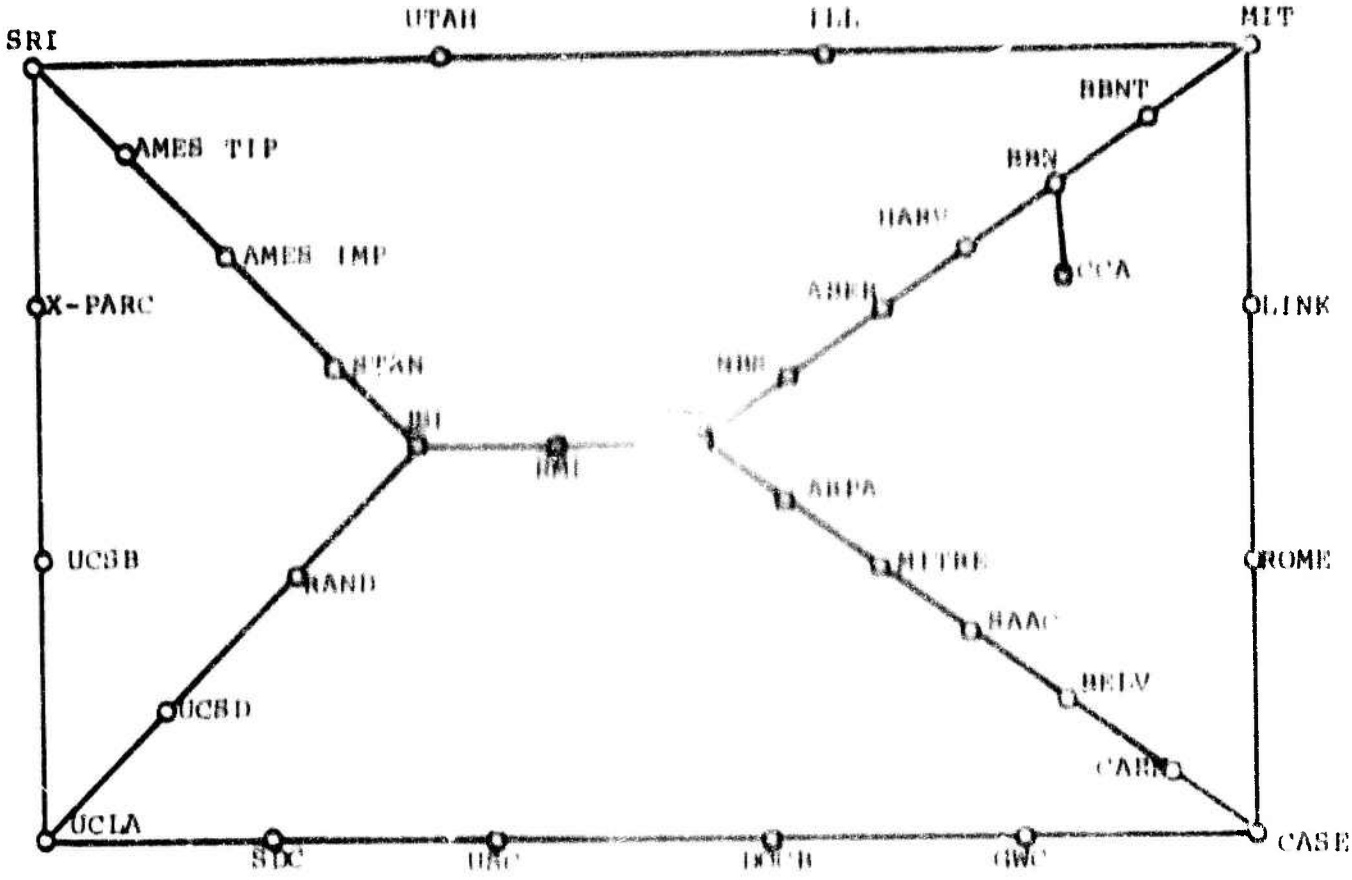

SRI
UTAH
ILL.
MIT
AMES TIP
AMES IMP
X-PARC
UCSB
UCSD
UCLA
SDC
RAND
BBNT
BBN
HARV
CCA
ABER
NBS
ARPA
MITRE
BAAC
BELV
CASE
ROME
LINK
GWC

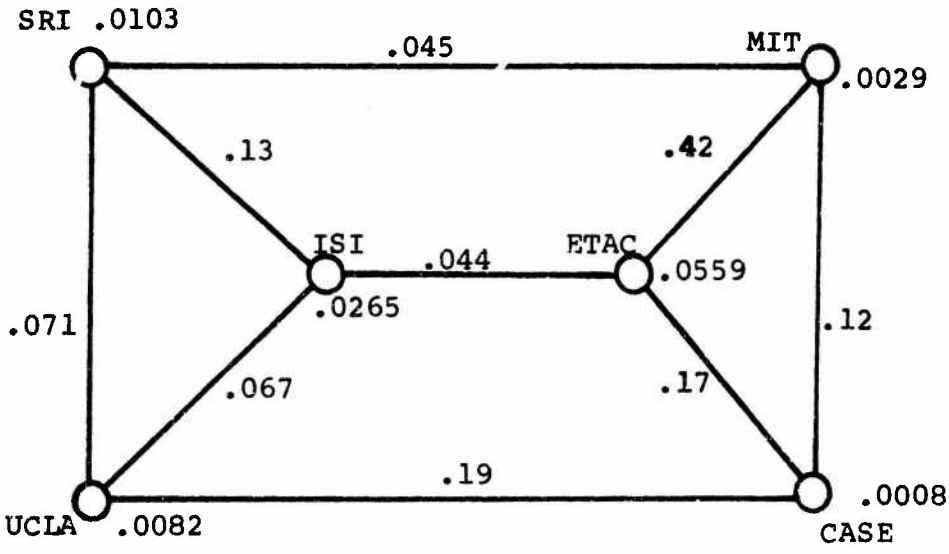
FIGURE 1.1

TABLE 1.1

4.8 Links Added	Additional Cost ($/Month)	Fraction of Node Pairs Disconnected	Probability of Network Disconnected
No links added	-	0.072	0.207
No links added (but pendant node inserted in MIT-ETAC chain)	-	0.080	0.185
CCA Pendant Node eliminated	-	0.068	0.120
(1)	435	0.061	0.101
(1) + (2)	860	0.053	0.076
(1) + (2) + (3)	1,295	0.052	0.057
(1) + (2) + (4)	1,250	0.051	0.056
(1) + (2) + (4) + (5)	1,495	0.050	0.051
(1) + (2) + (4) + (6)	2,790	0.051	0.049

Link labels:
(1): CCA-SAAC
(2): DOCB-UTAH
(3): ARPA-LINK
(4): ARPA-ROME
(5): AMES IMP-XPARK
(6): UCLA-MIT

the higher the probability that two elements fail in the same chain, thus producing a disconnected net. In order tc isolate the long chain contribution, a 32-node ARPANET, without the CCA pendant node, was first considered: the net disconnection probability was, in such a case, 0.12. Next, a "collapsed" version of the 32-node network was considered, in which each chain was replaced by a single link with failure probability equal to the probability of one or more failures in the original chain (see Figure 1.2). Obviously, in the collapsed network, the long chain contribution to the network disconnection probability has disappeared. For the collapsed network, the disconnection probability was 0.02. Therefore, more than 80% of the disconnections is produced by the chains!



"Collapsed" network with node failure rates and equivalent link failure rates.

FIGURE 1.2

3. INTRODUCTION OF LOW SPEED LINES

Having identified the main causes for poor network reliability, the 4.8 Kb links are introduced to eliminate these causes. For the 4.8 Kb links we assumed: line cost = 0.50 $/month/mile; modem cost = $120/end/month; link failure rate = .026.

The first link (see Table 1.1 and Figure 1.3) is introduced between CCA and SAAC; it makes the network 2-connected, and at the same time breaks two long chains (ETAC-MIT and ETAC-CASE). As a result, the probability of disconnection is reduced by half (see Table 1.1).

Additional improvement is obtained with the introduction of links (2), (3) and (4), which also break long chains. On the other hand, the introduction of link (5) or (6) does not produce significant improvement.

It is of interest to notice that in Table 1.1 the fraction of disconnected node pairs is only slightly reduced by the introduction of new links (from 0.07 to 0.05). This behavior is explained by the two following considerations. First, there is a lower bound $\cong$ 0.04 on the fraction of disconnected node pairs, which corresponds to the failure of source or destination node in the pair: the only way to improve such a bound is to make the nodes more reliable. Secondly, the disconnection of a pendant node, or of

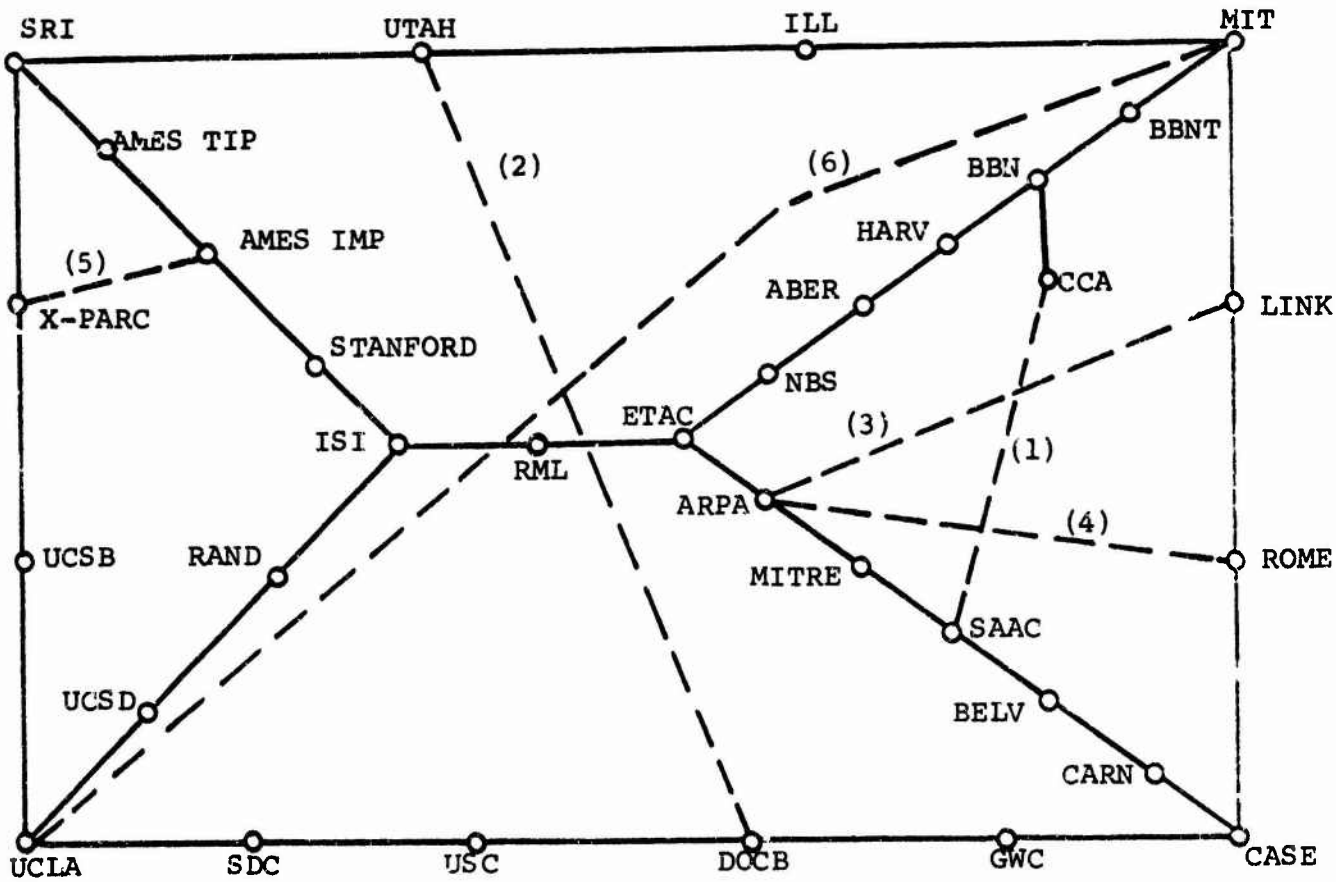


FIGURE 1.3

LOW SPEED LINE ADDITIONS FOR RELIABILITY ENHANCEMENT

a subset of nodes in the chain, produces a fraction of disconnected node pairs which is relatively small as compared to the total number of node pairs. Therefore, we cannot expect strong reduction of the fraction of disconnected node pairs only by eliminating pendant nodes and long chains.

From the results in Table 1.1, the most attractive solution from reliability-cost considerations seems to correspond to the introduction of links (1), (2) and (4). Therefore, such a solution was further investigated. First, we assumed uniform failure rates for nodes and links (equal to the respective averages: 0.021 for the links, 0.023 for the nodes). The results are:

Fraction of disconnected node pairs = .053

Probability of net disconnected = .054

Comparing these results to the values in Table 1.1, we notice that the overall performance is rather insensitive to deviations of node and link failure rates from the average values. In a second run, we assumed average failure rate for the nodes = 0.023, and zero failure rate for the links. The results are:

Fraction of disconnected node pairs = 0.047

Probability of network disconnected = .015

Here again we notice that by making the links perfectly reliable we considerably reduce the probability of network disconnected,

but only slightly decrease the fraction of disconnected node pairs, for the reasons previously exposed.

As a conclusion, the introduction of links (1), (2) and (4) seems to provide the best reliability-cost effectiveness for the current network. If the performance so obtained is still not satisfactory, substantial improvement to node reliability has to be made.

4. BY-PASS SWITCH ANALYSIS

4.1 Evaluation of P_{nc} and F_{nc}

The NAC computer program for reliability analysis computes P_{nc} and F_{nc}, given the network topology and the failure rates for nodes and arcs. Such a program, at least in its most general version, cannot be directly applied to networks with by-pass switches. The network with switches was therefore transformed in an equivalent "switchless" network, suitable for the reliability program.

The following transformations were performed:

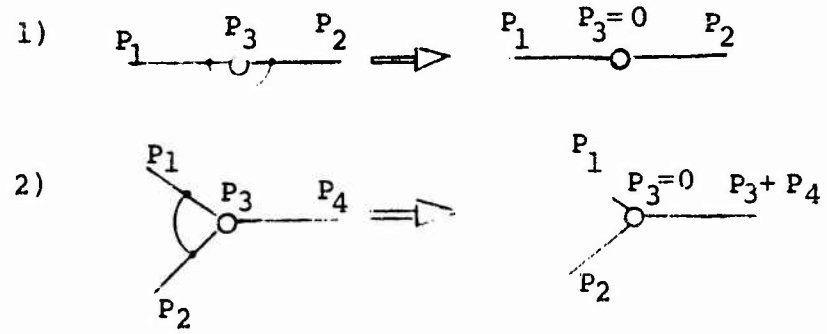


FIGURE 1.4

Essentially, by-passed nodes were made perfectly reliable (at the expense of the third link in the case of 3-degree nodes.)

Let P'_{nc} and F'_{nc} be the reliability parameters evaluated on the switchless network, and let P_{nc} and F_{nc} be the parameters

of the network with switches. It can be shown that, with excellent approximation:

$$\begin{cases} P_{nc} = P'_{nc} \\ F_{nc} = F'_{nc} + \frac{2}{N} \sum_{s \in S} P_s \end{cases}$$

where: N = total number of nodes

S = Set of nodes equipped with by-pass switches.

The correction to F_{nc} corresponds to the node pairs that fail when the by-passed node fails.

4.2 Average Throughput

The exact evaluation of the average throughput is very time consuming, as it requires the throughput computation for a large variety of network configurations, resulting from all possible component failures. An exact analysis for a 23-node ARPANET configuration was carried out assuming 2% down time for nodes and links, and is described in NAC-ARPA Report #4. The average throughput resulted to be 20% less than the maximum throughput.

For the purposes of this study it is sufficient to evaluate the amount of throughput which is lost because of a node failure, and the amount that can be recovered with a switch. As a first

approximation, the loss is given by the sum of: (1) traffic directed to the node; (2) traffic originated in the node; (3) traffic transiting through the node. Similarly the amount of throughput recovered equals the traffic transiting, before the failure, on the two links connected by the by-pass. The above terms can be easily computed from the traffic requirement matrix and from the channel data rates during normal network operation (no failures).

5. OPTIMAL ALLOCATION OF BY-PASS SWITCHES

We want to install the by-pass switches at strategic locations, in order to maximize the improvement of F_{nc}, P_{nc} and average throughput. It is computationally prohibitive to try all possible combinations. Instead, some reasonable criteria were developed, in order to identify the most critical nodes. The following criteria were considered:

a) Most Unreliable Nodes

The nodes were ranked in order of decreasing failure rate. Switches were installed at the most unreliable nodes.

b) Most Vital Nodes

The vitality of a node (or, better, of a switch installed at a node) is defined as the amount of throughput that the by-pass can recover after the node failure. The nodes were ranked in order of decreasing vitality, and switches were installed at the most vital nodes. This strategy is probably the most efficient for the case of enemy attacks to nodes; it is, in general, unsatisfactory if nodes fail with given probabilities different from node to node.

c) Most Critical Nodes

Node criticality is defined as node vitality times node failure rate. Thus, the criticality of a node is the

average throughput improvement obtained by installing a by-pass switch at that node. Again, the nodes were ranked in order of decreasing criticality, and switches were installed at the most critical nodes.

d) Most Critical 3-Degree Nodes

Considering that 3-degree node failures have a more severe impact on network reliability than 2-degree node failures, by-pass switches were installed at the most critical 3-degree nodes.

For each of the first three criteria, three "by-passed" configurations, with 5, 10, and 15 switches respectively, were considered. For the last criterion, only one configuration with 4 switches was analyzed. The results are presented in the next section.

6. EXPERIMENTAL RESULTS

Figure 1.5 shows the 32 node ARPANET configuration considered in this analysis.* Link and node probabilities were obtained from recent NCC cummulative statistics (See Appendix A). Various possible allocations of switches to nodes have been examined, according to the criteria mentioned in Section 3. The reliability results follow:

6.1 Evaluation of F_{nc} and P_{nc}

First, the following configurations without switches were analyzed:

a) Original net, no switches.

b) Original net, perfectly reliable nodes

Next, the following switch allocations were evaluated:

c) switches at all nodes

d.1) switches at 5 most unreliable nodes (See Table 1.2)

d.2) switches at 10 most unreliable nodes

d.3) switches at 15 most unreliable nodes

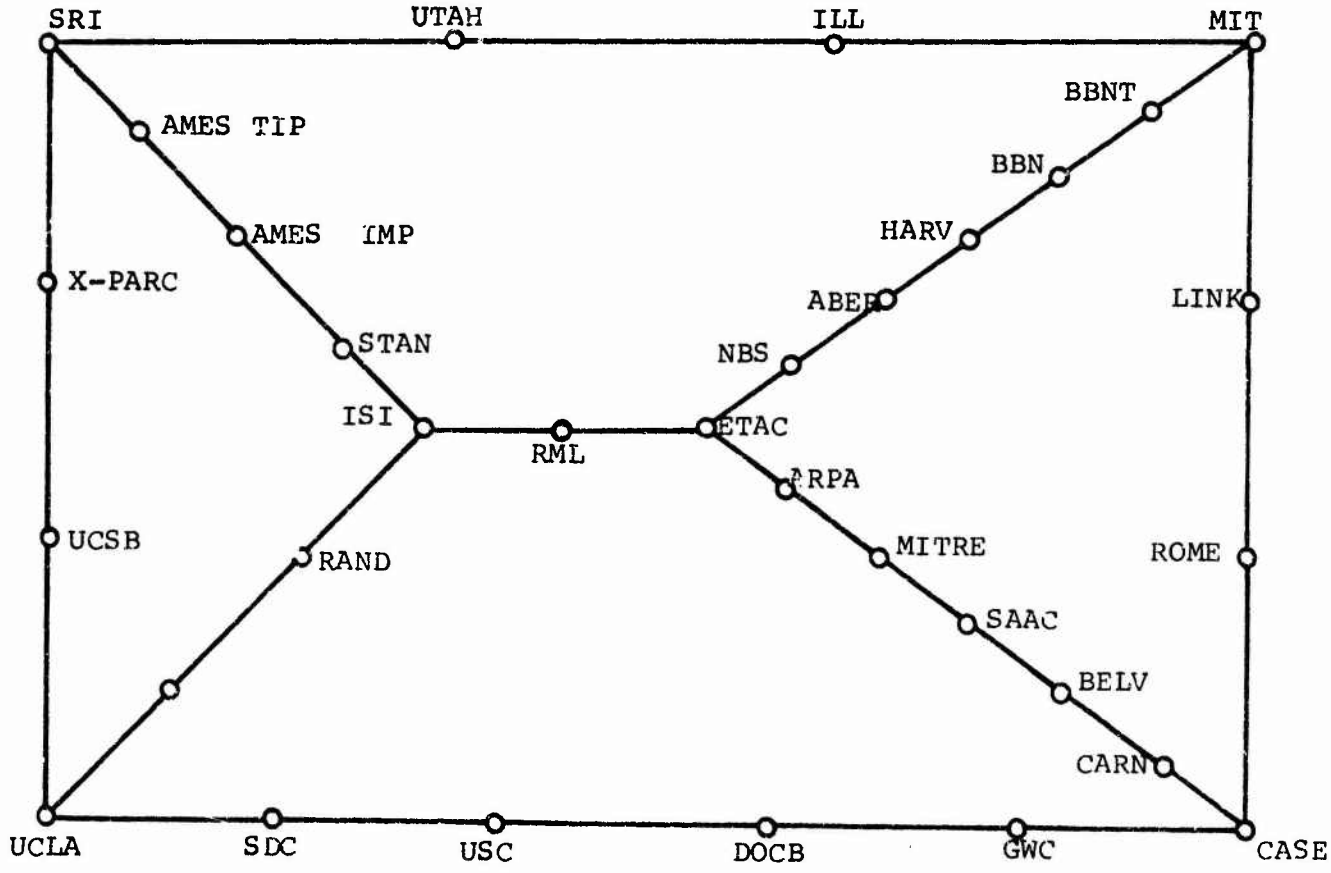
e.1) switches at 5 most vital nodes (See Table 1.2)

e.2) switches at 10 most vital nodes

e.3) switches at 15 most vital nodes

f.1) switches at 5 most critical nodes (See Table 1.2)

*Notice that the pendant node CCA, present in the early '73 ARPANET configuration, was not considered in this analysis because its very large contribution to P_{nc} ($\cong$ 10%)would have covered all relative improvements obtained with the switches.



ARPANET CONFIGURATION IN EARLY 1973
(without CCA pendant node)

FIGURE 1.5

TABLE 1.2

SWITCH ALLOCATION CRITERIA

Most unreliable* Nodes	Failure Rate	Most Vital Nodes	Vitality (Kbs)
GWC	.093	UTAH	98
ETAC	.056	ISI	96
BELV	.049	TINK	96
ROME	.054	STAN	95
MITRE	.038	SRI	95
ARPA	.037	ILL	93
ABRD	.037	AMES IMP	92
DOCB	.035	AMES TIP	92
ISI	.025	MIT	92
AMES TIP	.049	BBN TIP	92
AMES IMP	.044	ETAC	85
NBS	.023	BBN	75
TINK	.018	NBS	65
SRI	.010	ABER	60
USC	.025	HARV	60

* The following three factors were considered for node reliability: failure rate, node in long chain, node of degree three.

TABLE 1.2(Concluded)

Most Critical Nodes	Criticality (Kbs)
AMES TIP	4.5
AMES IMP	4.0
ETAC	3.9
GWC	2.5
ISI	2.4
ABRD	2.2
ROME	2.0
LL	2.0
TINK	1.6
ARPA	1.5
RAND	1.4
NBS	1.4
BELV	1.1
ILL	1.0

TABLE 1.3

RELIABILITY VS. SWITCH ALLOCATION

NUMBER AND LOCATION OF BY-PASS SWITCHES	FRACTION OF NON-COMMUNICATING NODE PAIRS F_{nc}	PROBABILITY OF NETWORK DISCONNECTED P_{nc}
Original net, without switches	.062	.118
Original net, with perfectly reliable nodes.	.006	.041
Switches at all nodes	.053	.046
Switches installed at unreliable nodes:		
5 switches	.059	.081
10 switches	.057	.067
15 switches	.057	.061
Switches installed at vital nodes:		
5 switches	.061	.108
10 switches	.059	.100
15 switches	.058	.076
Switches installed at critical nodes:		
5 switches	.060	.087
10 switches	.058	.070
15 switches	.056	.061
4 switches installed at the most critical 3-degree nodes	.059	.089

f.2) switches at 10 most critical nodes

f.3) switches at 15 most critical nodes

g) switches at 4 most critical 3 degree nodes.*

The values of F_{nc} and P_{nc} for the above configurations are given in Table 1.3. Upper and lower bounds on F_{nc} and P_{nc} are provided by Case a)and Case c). Reliability and criticality criteria seem to be the most effective in terms of reliability improvement. Among the two, the criticality criterion is probably the most desirable as it also provides the best average throughput.

6.2 Average Throughput

The maximum throughput for the original network configuration, with no failures, is 422 Kbs. The average throughput is approximately 25% less, i.e. 315 Kbs. The amount of average throughput recovered by switches equals the sum of the criticality values of the nodes where switches were installed. For example, if 5 switches are installed at the most critical nodes, the predicted average throughput improvement is 17.3 Kbs (approximately 5%).

In order to evaluate the cost-throughput effectiveness of the switches, one must recall that the incremental cost for the present ARPANET configuration is approximately 150$/Kbs x month, assuming that average throughput is increased by purchasing additional 50 Kbs channels. Considering that the approximate

*The 4 nodes are: ETAC, ISI, SRI and UCLA.

cost of a switch is 200 $/month, the installation of 5 switches in the most critical locations buys 17.3 Kbs for $1,000, as compared to the $2,500 required if 50 Kbs channels were added.

In order to verify experimentally the amount of throughput that a switch can in fact recover, the throughput with and without switches was computed for two node failures. The results follow

Case A
- AMES IMP down, no switch, throughput = 288 Kbs
- AMES IMP down, switch, throughput = 436 Kbs
- Throughput recovered by switch = 148 Kbs

Case B
- ETAC down, no switch, throughput = 276 Kbs
- ETAC down, switch, throughput = 412 Kbs
- Throughput recovered by switch = 136 Kbs

The values of recovered throughput are larger than those estimated in Table 1.2. The estimate therefore seems conservative, and the installation of switches appears even more attractive.

7. DISCUSSION

Recall that the insertion of three 4.8 Kbs lines gave the following results:

incremental cost: 1,250 $/month

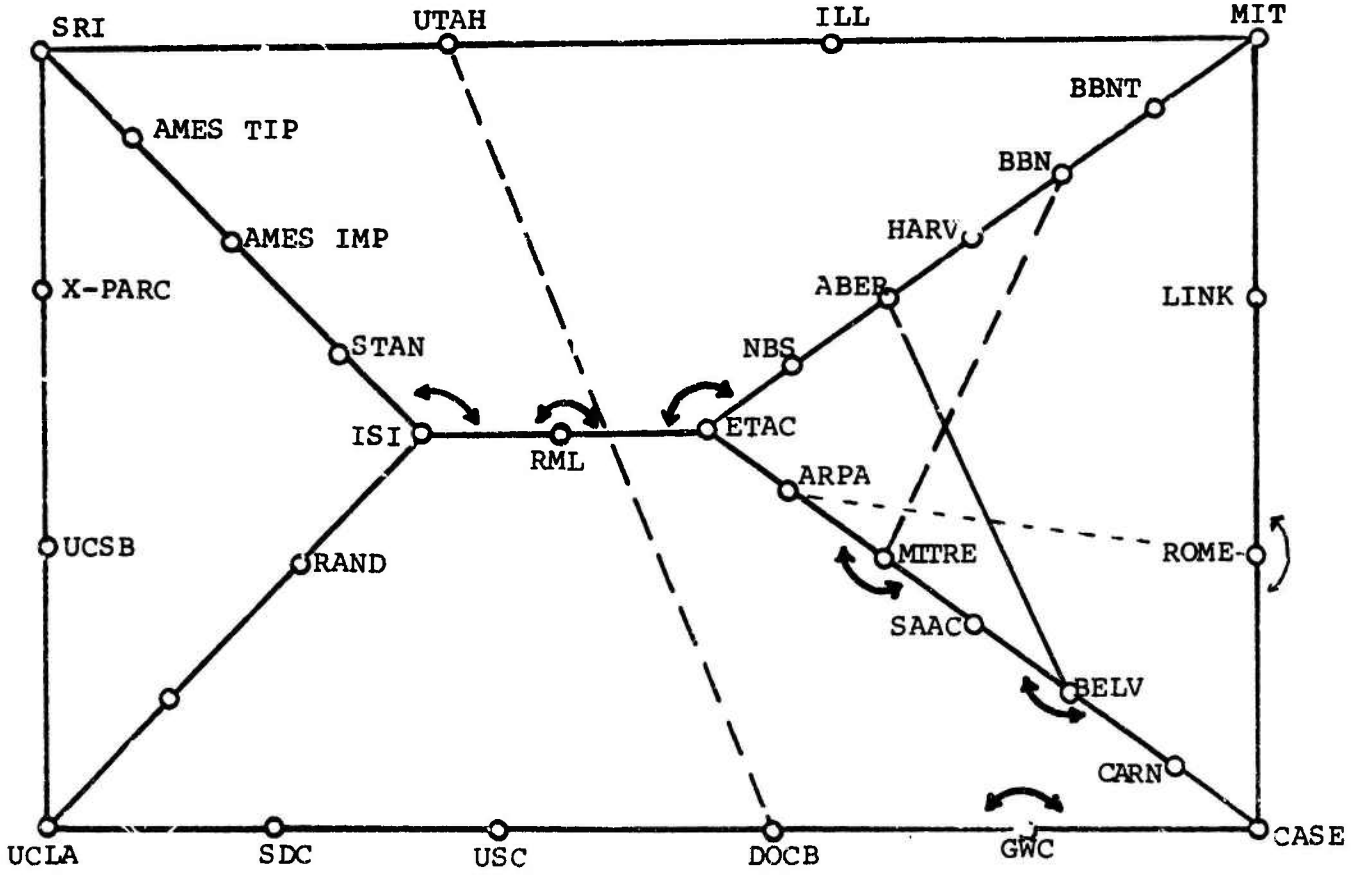
F_{nc} : .051

P_{nc} : .056

These values are considerably better than those obtained, for the same cost, with the switches. Therefore, if improvement of F_{nc} and P_{nc} is the only concern, 4.8 Kbs lines, rather than switches, should be installed.

On the other hand, 4.8 Kbs. channels provide neglible throughput improvement. In order to obtain the same improvement as with the switches, 50 Kbs channels must be used. However, the cost of the channels would be about 3 times higher than the cost of the switches. Therefore, if the main concern is throughput, switches at critical nodes should be installed.

In practical cases, both average throughput and connection probability are important. Therefore, a third strategy consists of introducing both 4.8 Kbs channels and bypass switches. In the 32-node ARPANET example, a proper combination of three 4.8 Kbs channels and seven switches (see Figure 1.6) gave the following results:



PROPOSED ARPANET CONFIGURATION FOR
RELIABILITY ENHANCEMENT WITH SEVEN
SWITCHES AND THREE 4.8 Kbs LINES

FIGURE 1.6

—— 50 Kbs

---- 4.8 Kbs

↷ Switch arrangement

F_{nc} = .027

F_{nc} = .051

Average throughput improvement = 14.5 Kbs

The incremental cost of such a configuration is approximately 2,650 $/month.

Considering the cost-reliability effectiveness of the above mentioned solutions, the third solution seems to be the most appropriate for the present ARPANET configuration.

8. ACCESS TO NETWORK RESOURCES

One of the primary goals of ARPANET is to provide access for all network users to network resources available at various computer sites. An important reliability measure is therefore the average fraction F_i of nodes which cannot access resources at node i, because of network component failures.

Among the resources presently available on ARPANET is the ILLIAC IV computer and the various PDP10 computers that offer TENEX System service. The values of F_i for all such resources have been computed for the ARPANET configuration of Figure 1.1. using the measured component failure rates. The results are shown in Table 1.4.

The fraction of nodes not communicating with at least one TENEX System was also evaluated. The result was:

$$F_{TENEX} = .026$$

Notice that a node might not be able to access a resource for one of the following reasons: (1) the node itself is down; (2) the resource node (or nodes) is down; (3) node and resource belong to disconnected components. The first two contributions provide the following lower bound on F_S. Let S be the set of nodes which offer a given resource, and let p_i be the failure rate for node i. We have:

TABLE 1.4

AVERAGE FRACTION OF NODES NOT COMMUNICATING WITH A GIVEN RESOURCE

Resource	Site	F_i
ILLIAC IV	AMES IMP	.075
TENEX	UTAH	.042
"	CCA	.14
"	SRI	.044
"	ISI	.063
"	CASE	.041
"	BBN	.062
"	AMES IMP	.075

$$F_S \geq \frac{\sum_{i \in S} p_i}{N-|S|} + \sum_{k \in S} p_k$$

Considering that the average node failure rate for the present ARPANET is 2.5×10^{-2}, we notice that many of the above results are close to the lower bound.

The installation of switches and 4.8 Kbs lines, is discussed earlier in this section, improves network connection probability and therefore reduces F_i. In particular, if 7 switches and 3 x 4.8 Kbs channels are installed, as shown in Figure 1.6, the fraction of nodes not communicating with ILLIAC IV becomes:

$$F = .065$$

The improvement, however, is not substantial, due to the fact that the value of F for the original network was already close to the lower bound.

Chapter 1--Appendix A

TABLE A-1

CUMULATIVE LINE OUTAGE SUMMARY
(JAN. 1971 THROUGH DEC. 1972)

LINE #	ID	FROM-TO	DAYS	DOWN TIMES	TOT TIME	PCT
34	6GW114CM	ABRDN-NBS	97	9	685:01	29.43
9	49GW175	ABER-BELV	69	5	211:54	12.80
31	7GW2176	CCA-BBN	67	4	164:14	10.21
39	TP553	HWAI-AMES2	11	2	15:54	6.02
18*	GW7518	BURR-CARN	201	29	281:59	5.85
9*	3GW3305	MCCL-SRI	178	20	229:36	5.37
2	07GW2074	BBN-MIT	731	98	915:04	5.22
1*	GW7533	TINK-BBN	166	28	164:40	4.13
15*	GW7509	MITRE-CARN	266	22	217:20	3.40
6	9GW4301	UCSD-UCLA	47	5	36:44	3.26
16*	GW7503	CASE-LINC	372	38	278:41	3.12
19*	GW7553	NBS-HVD	221	14	136:51	2.58
25	GW7532	ISI-TINK	298	19	173:40	2.43
26	07GW3007	USC-SDC	278	24	160:15	2.40
19*	GW7508	BURR-HVD	363	23	204:14	2.34
3	GW7512	ILL-MIT	713	75	347:28	2.03
4*	GW7502	SDC-UTAH	453	16	202:57	1.87
24	75GW6001	AMES -AMES2	294	2	128:37	1.82
1*	GW7511	RAND-BBN	433	54	183:14	1.76
10*	GW7531	MCCL-UTAH	178	23	70:37	1.65
23	14GW165	ETAC-NBS	397	22	144:22	1.52
13	07GW4000	HVD-BBN	731	17	256:23	1.46
19	GW7546	ABRDN-HVD	97	10	33:23	1.43
7	7GW1	UCSB-UCLA	731	27	239:52	1.37
28	37GW0171	BELV-SAAC	214	29	61:19	1.19

TABLE A-I (Continued)

LINE #	ID	FROM-TO	DAYS	DOWN TIMES	DOWN TOT TIME	DOWN PCT
21*	14GW161	ETAC-MITRE	198	9	55:06	1.16
16	GW7517	ROME-LINC	324	36	87:14	1.12
17	GW7504	CARN-CASE	692	45	165:59	1.00
8*	1GW547	UCSB-SRI	669	43	158:59	.87
12	1GW1095	ISI-STAN	731	31	147:20	.84
21*	GW7535	MITRE-BURR	83	5	15:54	.80
22	GW7516	ROME-CASE	324	19	59:02	.76
6*	PL778346	RAND-UCLA	642	8	115:11	.75
18	GW7543	BELV-CARN	214	29	36:32	.71
33	34GW0172	ARPA-ETAC	199	2	32:12	.67
14	08GW2093	LINC-MIT	723	20	116:57	.67
9*	1GW548	SRI-UCLA	349	25	55:39	.66
31*	2GW1498	NOAA-UTAH	74	8	11:22	.64
15	GW7513	ILL-UTAH	707	57	108:03	.64
10*	GW7501	UTAH-SRI	444	26	55:17	.52
27	GW7523	GWC-CASE	272	18	32:58	.51
11	1GW2568	AMES1-SRI	515	11	52:41	.43
32	GW7524	NOAA-GWC	214	12	20:22	.40
4*	GW7537	USC-UTAH	140	5	13:05	.39
35	1GW6015	XEROX-UCSB	62	3	5:38	.38
24*	GW7534	NBS-BURR	79	6	6:36	.35
38	9GW4300	UCSD-RAND	48	7	3:56	.34
4	GW7552	NOAA-USC	138	9	11:11	.34
20	1GW2547	AMES2-STAN	516	9	39:09	.32
1	GW7549	TINK-ETAC	84	3	3:00	.15

TABLE A-I (Concluded)

LINE #	ID	FROM-TO	DAYS	DOWN TIMES	DOWN TOT TIME	DOWN PCT
29	37GW0166	SAAC-MITRE	214	3	5:36	.11
5*	PL778347	SDC-RAND	570	2	12:19	.09
21	34GW0173	ARPA-MITRE	199	5	4:17	.09
5	198-0080	SDC-UCLA	161	3	2:39	.07
11*	77GW1096	STAN-SRI	216	3	3:29	.07
36	198-0143	ISI-RAND	83	2	0:41	.03
10	GW7571	UTAH-SRI	109	1	0:16	.01
6*	198-0079	IIS-UCLA	42	1	0:06	.01

ALL OTHER LINES HAD ZERO OUTAGES DURING THIS PERIOD.

TOTALS OVER ALL LINES	1081	6963:05	
AVERAGE OUTAGE LENGTH OVER ALL LINES		6:26	
AVERAGE % DOWN OVER ALL LINES			1.64

* NO LONGER IN SERVICE AT THE END OF THIS PERIOD.

TABLE A-II

CUMULATIVE IMP DOWN SUMMARY
(JUNE 1972 THROUGH DEC. 1972)

Site	Months In Service	# Times Down	Total Down Time	%
GWC	10	43 (29)	481:08 (446:33)	9.37 (8.69)
ETAC	13	54 (37)	287:21 (140:11)	5.59 (2.73)
Rome	11	29 (12)	278:27 (97:03)	5.42 (1.89)
Ames #1	16	57 (37)	254:26 (135:39)	4.95 (2.64)
Belvoir	7	19 (8)	254:25 (181:42)	4.95 (3.54)
Lincoln	24	28 (24)	244:51 (73:16)	4.77 (1.43)
Ames #2	22	13 (3)	228:10 (26:27)	4.44 (.51)
Mitre	16	24 (12)	195:25 (96:23)	3.80 (1.88)
ARPA	7	20 (13)	179:33 (165:50)	3.76 (3.47)
Aberdeen	4	18 (15)	85:49 (83:42)	3.72 (3.63)
NOAA	7	20 (9)	180:29 (113:45)	3.51 (2.21)
ISI	3	21 (10)	135:58 (115:57)	2.65 (2.38)
Xerox	3	8 (3)	56:24 (18:43)	2.53 (.84)
USC	10	65 (35)	127:20 (63:19)	2.48 (1.23)
Rand	24	16 (8)	123:30 (39:40)	2.40 (.77)
NBS	13	23 (9)	121:29 (58.17)	2.36 (1.13)
Tinker	10	27 (5)	93:16 (18:50)	1.81 (.37)
Illinois	22	14 (4)	58:19 (36:33)	1.13 (.71)
SRI	24	11 (7)	52:56 (49:15)	1.03 (.96)
UCLA	24	24 (15)	41:58 (24:23)	.82 (.47)
CCA	3	3 (3)	17:37 (17:37)	.81 (.81)
SAAC	7	20 (13)	36:57 (31:41)	.72 (.62)
UCSD	2	2 (2)	7:41 (7:41)	.71 (.71)
Stanford	24	13 (6)	29:19 (20:25)	.57 (.40)
Harvard	24	17 (12)	18:31 (13:53)	.36 (.27)

TABLE A-II (Concluded)

Site	Months In Service	# Times Down	Total Down Time	%
Utah	24	12 (5)	17:13 (7:52)	.33 (.15)
MIT	24	19 (10)	14:52 (3:28)	.29 (.07)
UCSB	24	9 (5)	13:58 (6:15)	.27 (.12)
BBN	24	16 (12)	13:17 (10:11)	.26 (.20)
Carnegie	23	12 (2)	12:11 (2:50)	.24 (.05)
SDC	24	9 (4)	4:06 (2:03)	.08 (.04)
Case	22	4 (2)	4:02 (0:02)	.08 (.00)
Hawaii	1	0 (0)	0:00 (0:00)	.00 (.00)

( ) Denotes IMP Hardware/Software Failures

Total Machine Hours	152448	(152448)
Total Down Time	3670:58	(2109:26)
Number of Times Down	670	(371)
Percent	2.42	(1.39)
MTBF	227.5342	(410.9111)
MTTR		(5:41)

2. ARPANET GROWTH, THROUGHPUT AND TRAFFIC SENSITIVITY

1. INTRODUCTION

During the reporting period, a number of new IMP and TIP locations were added to the ARPANET, and several others were proposed. The recommended topological modifications to accommodate various new locations are presented in Section 3. As new locations are introduced into the network, the projected capacity (throughput) has been systematically reduced. This induced a further study of the economic and growth characteristics of the ARPANET, i.e. the incremental costs required to increase the network's throughput. The results are detailed in Section 2.

These studies were performed by assuming a uniform traffic distribution; i.e., the traffic requirements between all node pairs are the same. Previous studies performed by NAC had shown that the degradation in performance caused by variations in traffic requirements is not great. Therefore, the uniform traffic distribution can be justified as a good assumption when the actual traffic pattern is unknown. (Details are given in the Final Technical Report to Contract DAHC 15-70-C-0120.) However, these studies were carried out without any use of the actual traffic pattern of the current ARPANET. Now that traffic measurements are available and certain traffic trends can be

observed, new studies were performed to test the validity of the uniform traffic distribution assumption. Section 4 shows the results of the study.

Throughout this chapter, throughput is defined to be the traffic in the network when the average single packet delay, from the originating IMP or TIP to the destination, is 0.2 seconds. The average information packet length is assumed to be 627 bits. This is a number estimated over three years ago. Since there is no better estimate, it is still being used in our analysis. However, in the current traffic environment, one can be quite certain that the average packet length is less than 627 bits. Therefore, the throughputs in packets per day derived in this chapter are conservative estimates.

Early this year (1973), BBN has modified the link control procedure so that node-to-node acknowledgments may be "piggy-packed" onto messages flowing in the reverse direction. This results in a throughput increase of about 17%. The throughputs shown in this report have taken this into account. Some items which had been calculated before this modification have been adjusted to reflect the new acknowledgment procedure.

2. ARPANET GROWTH

During this reporting period a total of nine new locations have been proposed and one has been deleted.

If the locations of all network nodes are known in advance, it is clearly most efficient to design the topological structure as a single global effort. However, in the ARPANET, as in most actual networks, node locations are added and modified on numerous occasions. On each such occasion, the topology could be completely reoptimized to determine a new set of link locations.

In practice, however, there is a long lead time between the ordering and the delivery of a link and major topological modifications cannot be made without substantial difficulty. It is therefore prudent to add or delete nodes with as little disturbance as possible to the basic network structure consistent with overall economic operation.

Figures 2.1(a), 2.1(b), and 2.1(c) show three proposed ARPANETs derived using the policy of minimum disturbance to the network for any topological modifications. They represent the proposed net in December of 1972, January of 1973 and May of 1974. Estimated line and modem costs and throughputs for the three nets are given in the last three lines of Table 2.1, respectively. Coordinates of node locations are shown in Table 2.2.

TABLE 2.1

NETWORK LINE COSTS

Number of Nodes	Yearly Line Cost (K$)	Throughput (Uniform Traffic) KBPS/Node	Throughput (Uniform Traffic) KPacket/Day/Node *	Line/Cost Node (K$)	Line/Co KPacket (cents
14	605	12.2	1690	43.2	7
15	659	12.5	1730	43.9	7
18	792	14.2	1970	44.0	6
21	825	12.4	1710	39.3	6
23	849	11.9	1640	36.9	6
24	860	11.1	1530	35.8	6
26	810	11.6	1600	31.2	5
30	859	10.1	1400	28.6	6
33	886	9.3	1290	26.8	6
39	1,016	8.7	1210	26.1	6
40	1,022	8.5	1180	25.6	6
39	1,012	8.0	1100	26.0	6
41	1,032	7.3	1000	25.2	7
46	1,119	6.4	875	24.3	7

*Based on 24 hr/day operation.

TABLE 2.2

	NAME	LATITUDE		LONGITUDE	
1.	UCLA	34	4	118	31
2.	SRI	37	22	122	10
3.	UCSB	34	30	119	45
4.	UTAH	40	40	111	50
5.	ISI	34	0	118	35
6.	CCA	42	30	71	20
7.	SDC	34	1	118	33
8.	MIT 1	42	30	71	12
9.	ILL	40	5	88	30
10.	HARV	42	30	71	15
11.	CMU	40	30	79	50
12.	ETAC	38	50	77	0
13.	SDAC	38	55	77	10
14.	LL	42	35	71	20
15.	CASE	41	30	81	45
16.	STAN	37	18	122	10
17.	MITRE	39	0	77	0
18.	DOCB	39	30	105	0
19.	LBL	37	50	122	17
20.	AFGWC	41	0	96	0
21.	RADC	43	15	75	25
22.	AMES IMP	37	17	122	2
23.	USC	34	0	118	21
24.	UAWC	41	32	90	34
25.	UAAC	38	40	90	15
26.	NBS	39	8	77	10
27.	ARPA	39	0	77	0
28.	BBN	42	30	71	15
29.	ABERDEEN	39	0	77	0
30.	BELVOIR	39	5	77	0
31.	AMES TIP	37	17	122	2
32.	X-PARC	37	18	122	10
33.	UCSD	32	40	117	10
34.	FNWC	36	30	121	55
35.	RML	28	15	80	34
36.	NYU	40	45	74	0
37.	RUTGERS	40	29	74	27
38.	LLL	37	38	121	45
39.	RAND	33	55	118	35
40.	MOFFETT	37	17	122	2
41.	MIT 2	42	30	71	12
42.	AEDC	35	10	86	10
43.	AFADC	30	20	87	20
44.	AFARL	39	45	84	12
45.	ANL	41	50	87	40
46.	AFWL	35	1	106	30

3. ARPANET THROUGHPUT/COST STUDY

The present traffic pattern in the ARPANET appears to be growing at a rate of 100% every ten months. Without lowering the traffic growth rate and/or increasing the network throughput, the network could become saturated within several years if this growth rate continues. This study is to estimate the increase in communication costs needed to accomodate additional traffic.

Several network topologies of different throughput levels are designed, with the lowest throughput represented by the projected 40 node early 1973 ARPANET and the highest one having a throughput over 200% greater. When a most cost-effective network topology is obtained for a certain throughput level, communication lines are added to form a new network topology with higher throughput. The optimization process allows only adding lines without deleting any from the network of lower throughput. It is so constrained because of the consideration that higher throughput can be obtained for any of the designs in this study without the need of altering any line of the network. The added lines are also restricted not to be connected to the sites that are projected but not yet in the ARPANET. This is so constrained because of the consideration that modifications suggested in this study can still be implemented even if some of the projected sites are not included in the ARPANET.

The study demonstrates that the percentage increase in communications costs is approximately one-half the percentage increase in throughput. Results are summarized in Figure 2.8 and Table 2.3. For each design listed in Table 2.3, detailed topological information is supplied on a separate figure. (One may notice that the projected throughput in this study for the early 1973 ARPANET is slightly higher than the one given before. This is because the link control procedure has been modified to eliminate the overhead traffic caused by acknowledgments.)

The throughput is obtained by requiring that the average response time for a packet to transit from its originating Host computer to its destination Host computer is no more than 0.2 seconds. In deriving the throughput, it is assumed that the traffic routing through the network is close to optimal and that the traffic generated from each site is the same as any other. However, the routing strategy used by the net may deviate from the best flow pattern and therefore may not be optimal. Furthermore, even though throughput is insensitive to traffic variations among different sites, it may vary a few percent as the traffic pattern varies. Due to these two considerations, it is advisable that the network not be operated normally with a traffic load over 90% of the throughput projected in Table 2.3.

It is suggested that a means to control traffic growth rate, (such as charging for packets) and/or an increase in network throughput should be in effect before the network traffic load reaches 90% of the projected throughput.

During the study, the possibility of using 230K lines and T-1 carriers have been explored. The investigation shows that within the throughput range studied, they cannot be economically utilized.

TABLE 2.3

THROUGHPUT VS. COST

Fig. #	KBPS/ Node	K-PKTS/ Hr/Node	M-PKTS/ Day/Node*	M-PKTS/Day (Entire Network)*	Line, Modem Costs	Additional Lines
2	7.3	42	1.0	40	1.023	-
3	9.4	54	1.3	52	1.168	(AMES IMP, MITRE)
4	11.8	67	1.6	64	1.356	(FNWC, UTAH) (UTAH, DOCB) (DOCB, HARV)
5	12.8	74	1.8	72	1.403	(AMES IMP, FNWC) (HARV, MITRE)
6	19.6	112	2.7	108	1.844	(UCSB, AMES TIP) (AMES TIP, LBL) (LBL, ILL) (ILL, ABERDEEN) (ABERDEEN, BBN) (BBN, LL), (LL, HARV) (LL, AFGWC), (AFGWC, SDC) (SDC, RAND), (RAND, UCSB)
7	21.7	124	3.0	120	2.053	(SRI, MIT, (ILL, BELVOIR)

*Based on 24 hours per day.

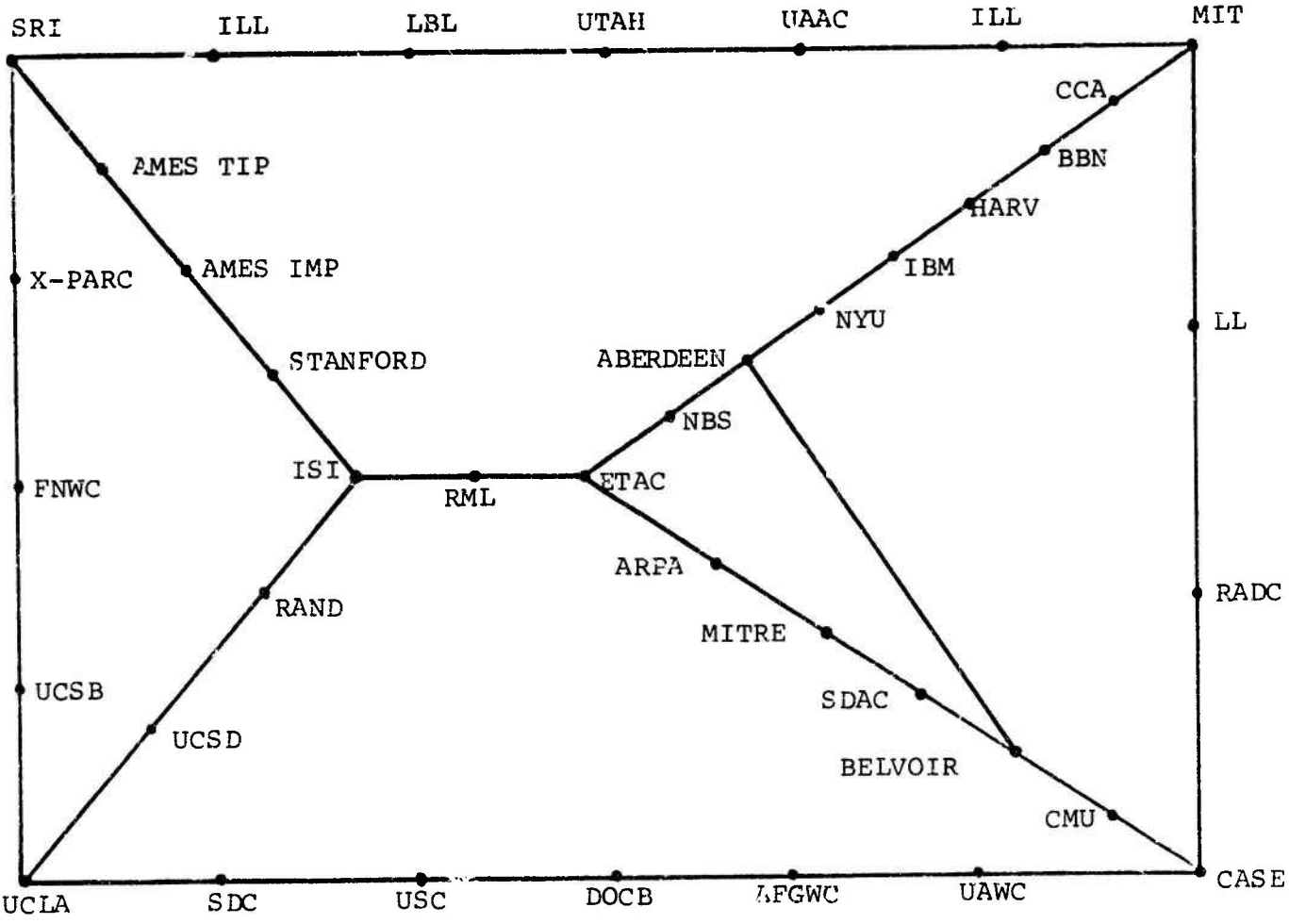

SRI
ILL
LBL
UTAH
UAAC
ILL
MIT
CCA
BBN
HARV
IBM
NYU
LL
AMES TIP
X-PARC
AMES IMP
STANFORD
ABERDEEN
NBS
ISI
ETAC
RML
FNWC
ARPA
RADC
RAND
MITRE
UCSB
SDAC
UCSD
BELVOIR
CMU
CASE
UCLA
SDC
USC
DOCB
AFGWC
UAWC

FIGURE 2.1(a)

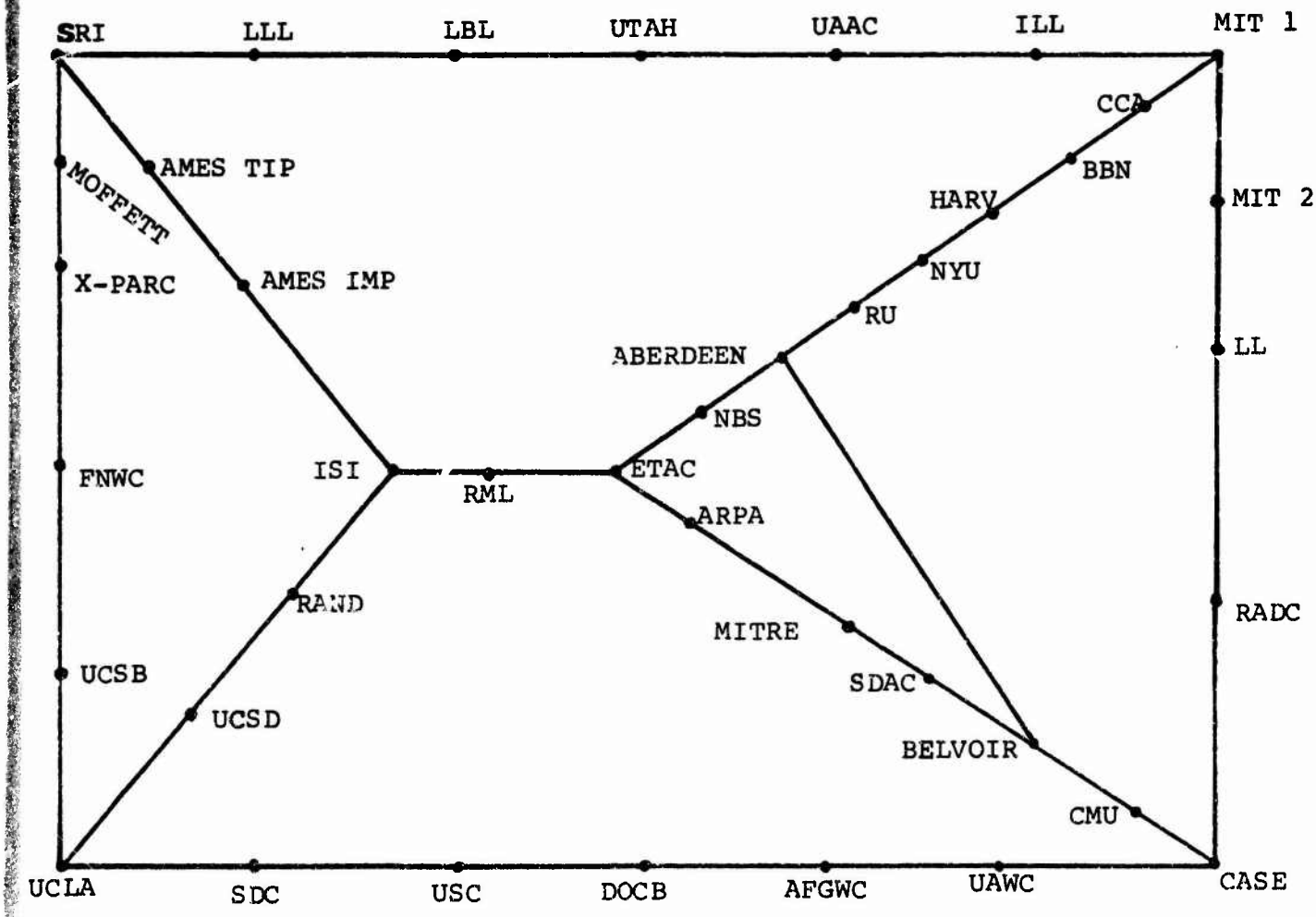

SRI
LLL
LBL
UTAH
UAAC
ILL
MIT 1
CCA
BBN
HARV
MIT 2
NYU
RU
LL
ABERDEEN
NBS
ETAC
ARPA
RADC
MITRE
SDAC
BELVOIR
CMU
CASE
AMES TIP
MOFFETT
X-PARC
AMES IMP
FNWC
ISI
RML
RAND
UCSB
UCSD
UCLA
SDC
USC
DOCB
AFGWC
UAWC

FIGURE 2.1(b)

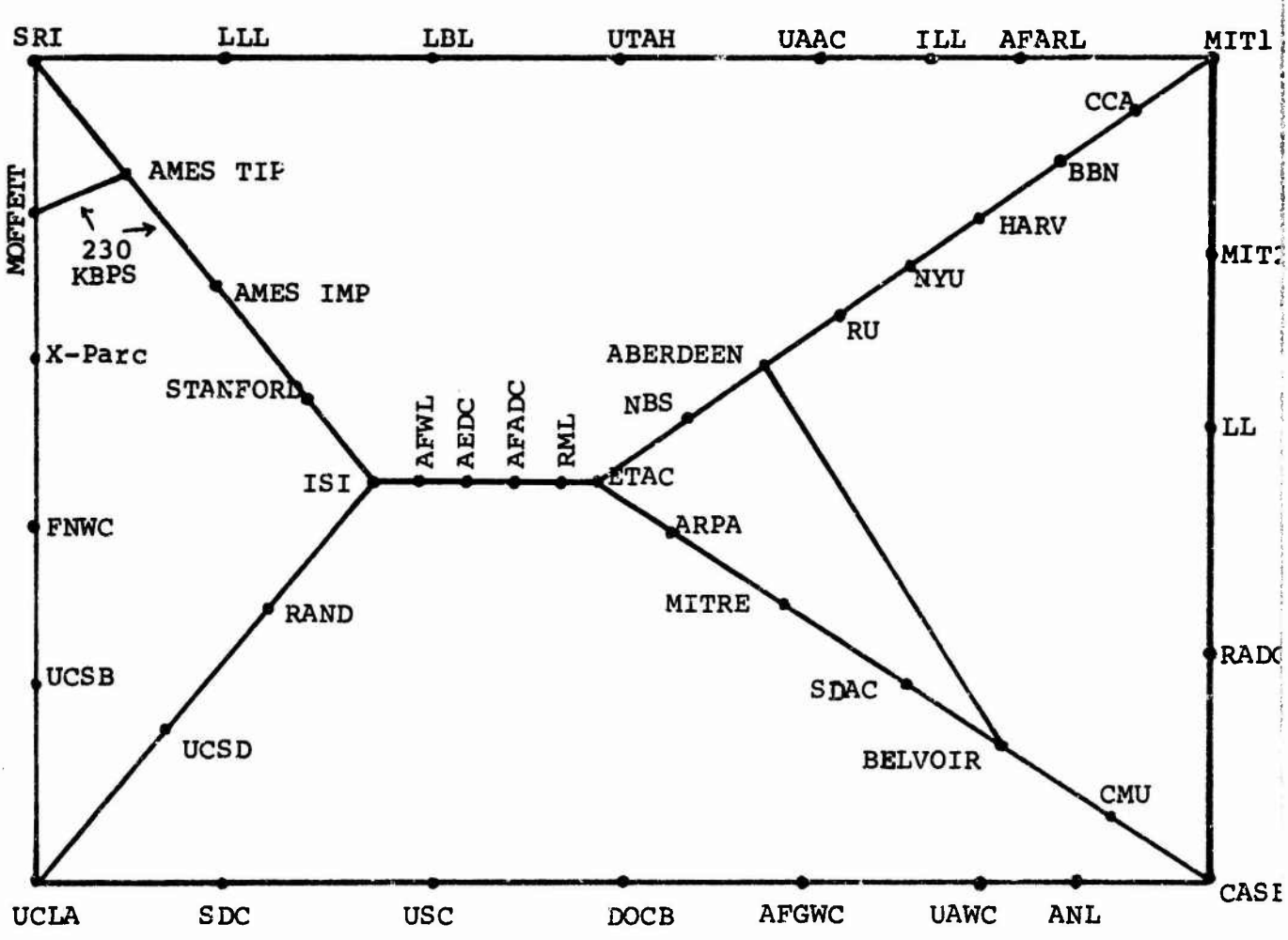

SRI
LLL
LBL
UTAH
UAAC
ILL
AFARL
MIT1
CCA
BBN
HARV
MIT2
NYU
RU
ABERDEEN
NBS
LL
ETAC
ARPA
MITRE
RADC
SDAC
BELVOIR
CMU
CASE
MOFFETT
AMES TIP
230
KBPS
AMES IMP
X-Parc
STANFORD
AFWL
AEDC
AFADC
RML
ISI
FNWC
RAND
UCSB
UCSD
UCLA
SDC
USC
DOCB
AFGWC
UAWC
ANL

FIGURE 2.1(c)

FIGURE 2.2

LINE AND MODEM COSTS: 1.023 Million Dollars

THROUGHPUT: 7.3 KBPS/Node
1.0 M-PKTS/Day/Node
40 M-PKTS/Day

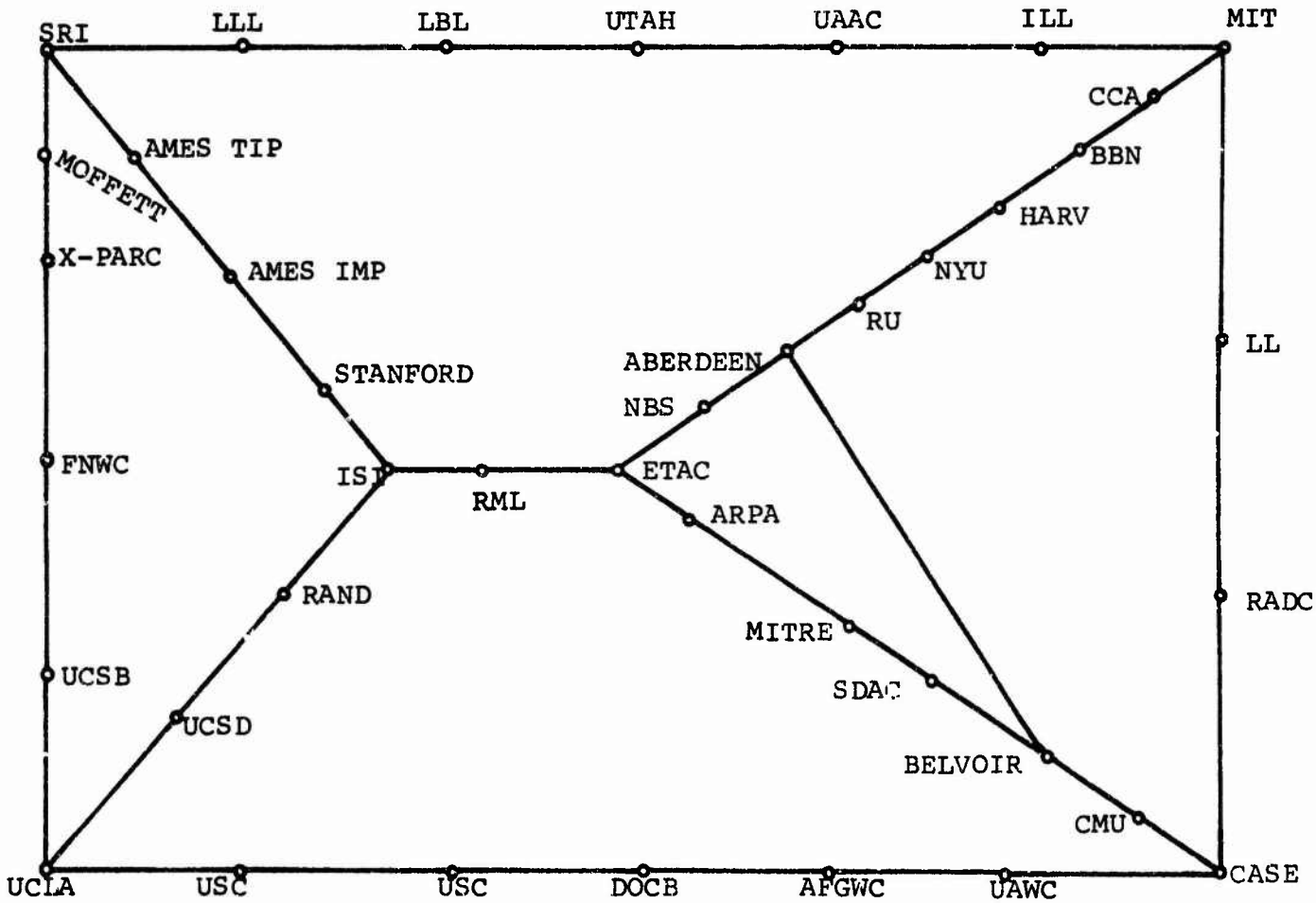


FIGURE 2.3

LINE AND MODEM COSTS: 1.168 Million Dollars

THROUGHPUT: 9.4 KBPS/Node
1.3 M-PKTS/Day/Node
52 M-PKTS/Day

LINE ADDED: AMES IMP-MITRE

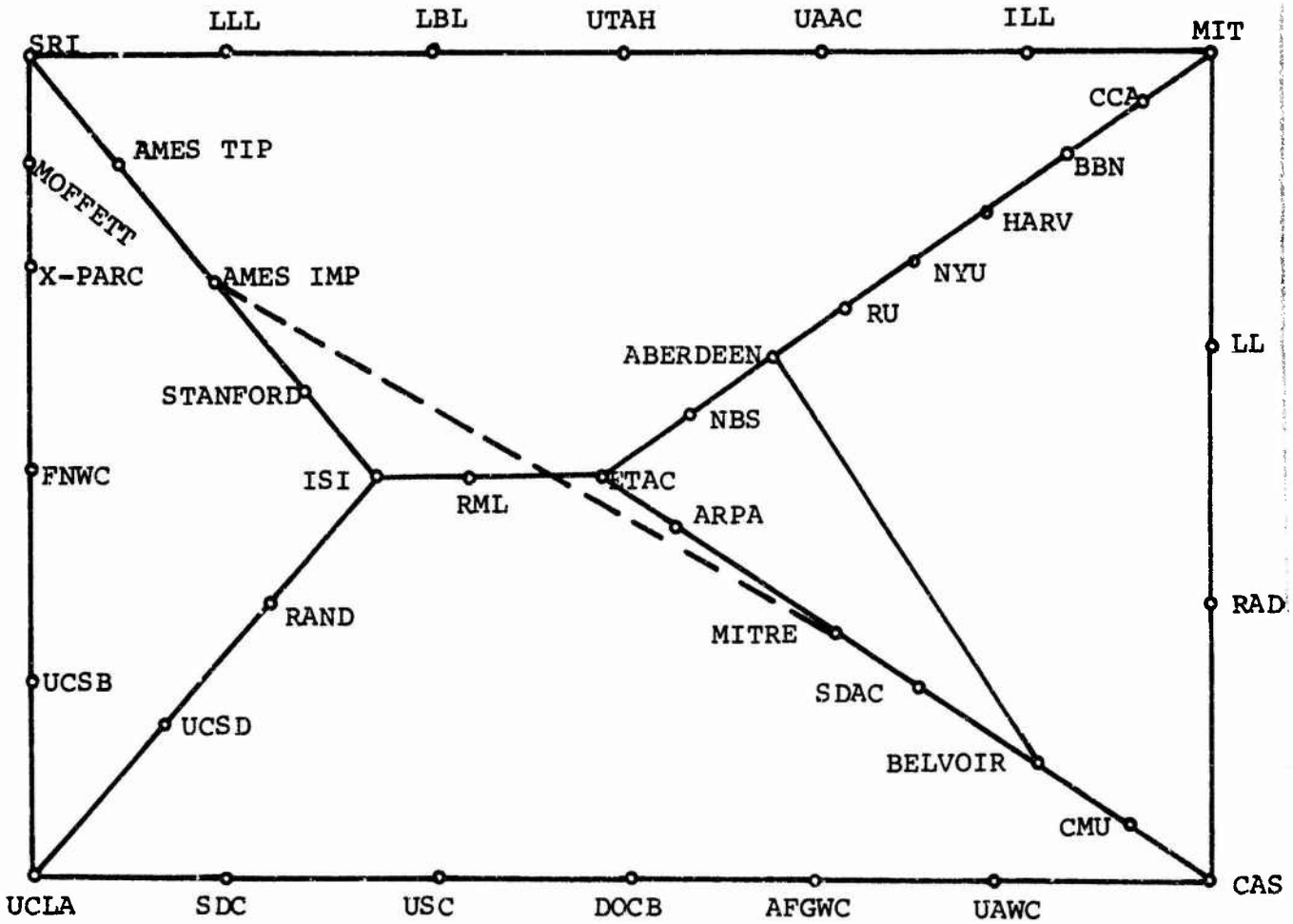


FIGURE 2.4

LINE AND MODEM COST: 1.356 Million dollars

THROUGHPUT: 11.7 KBPS/Node
1.6 M-PKTS/Day/Node
64 M-PKTS/Day

LINES ADDED: FNWC-UTAH, UTAH-DOCB, DOCB-HARV

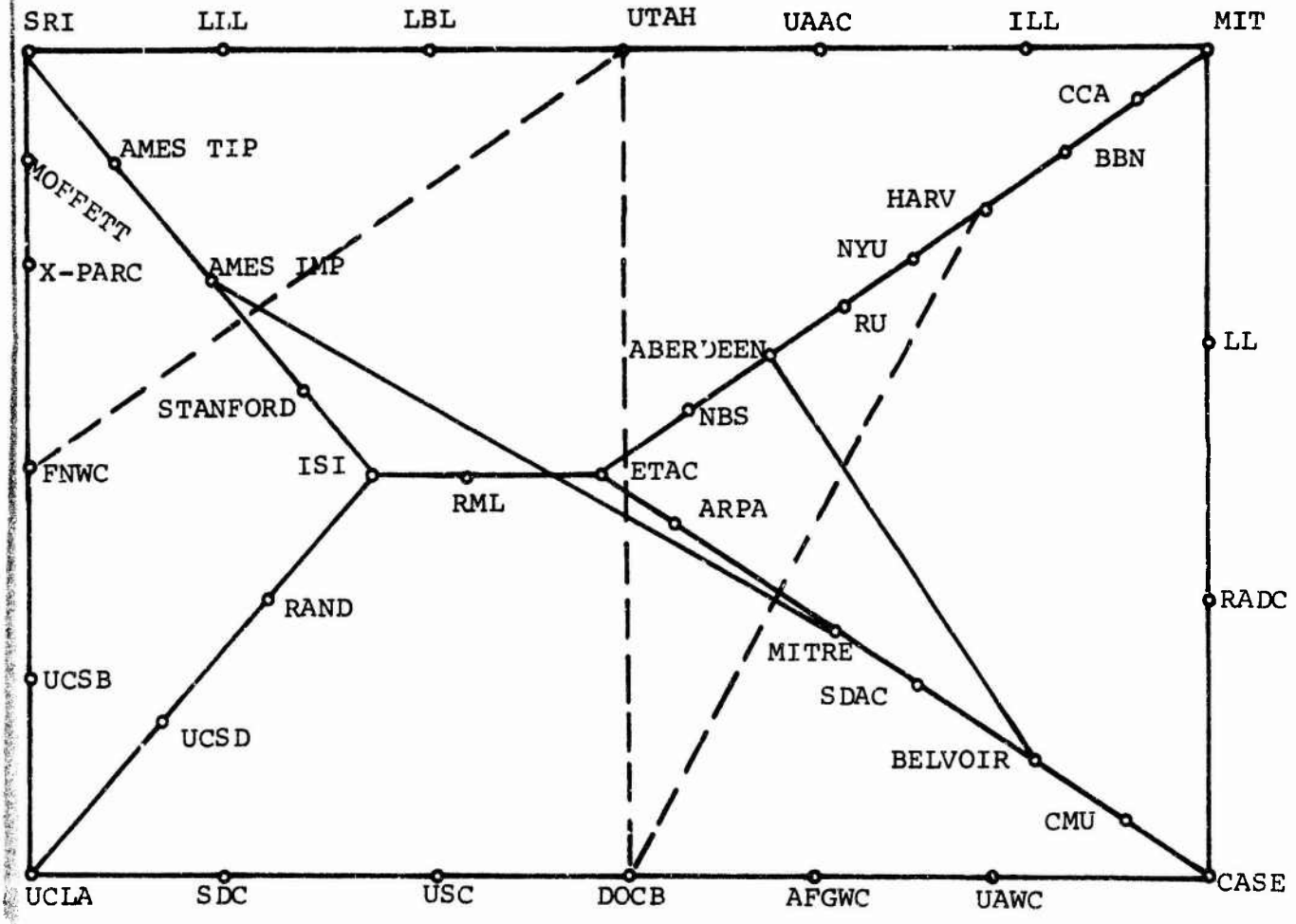


FIGURE 2.5

LINE AND MODEM COSTS: 1.403 Million Dollars

THROUGHPUT: 12.8 KBPS/Node
1.8 M-PKTS/Day/Node
72 M-PKTS/Day

LINES ADDED: FNWC-AMES IMP
MITRE-HARV

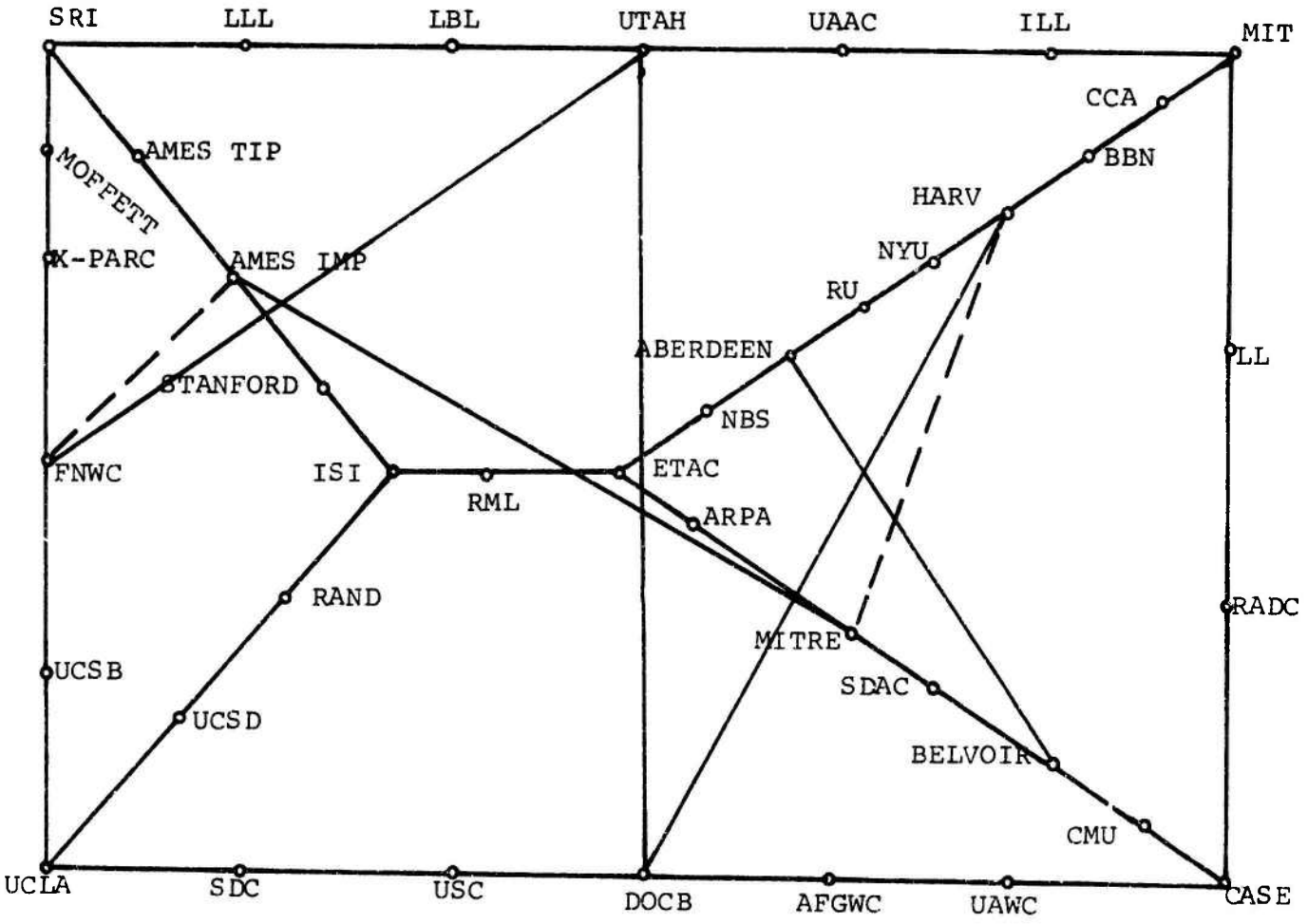


FIGURE 2.6

LINE AND MODEM COSTS: 1.844 Million Dollars

THROUGHPUT: 19.6 KBPS/Node
2.7 M-PKTS/Day/Node
108 M-PKTS/Node

LINES ADDED: UCSB-AMES TIP, AMES TIP-LBL, LBL-ILL, ILL-ABERDEEN, ABERDEEN-BBN, BBN-LL LL-HARV, LL-AFGWC AFGWC-SDC, SDC-RAND, RAND-UCSB

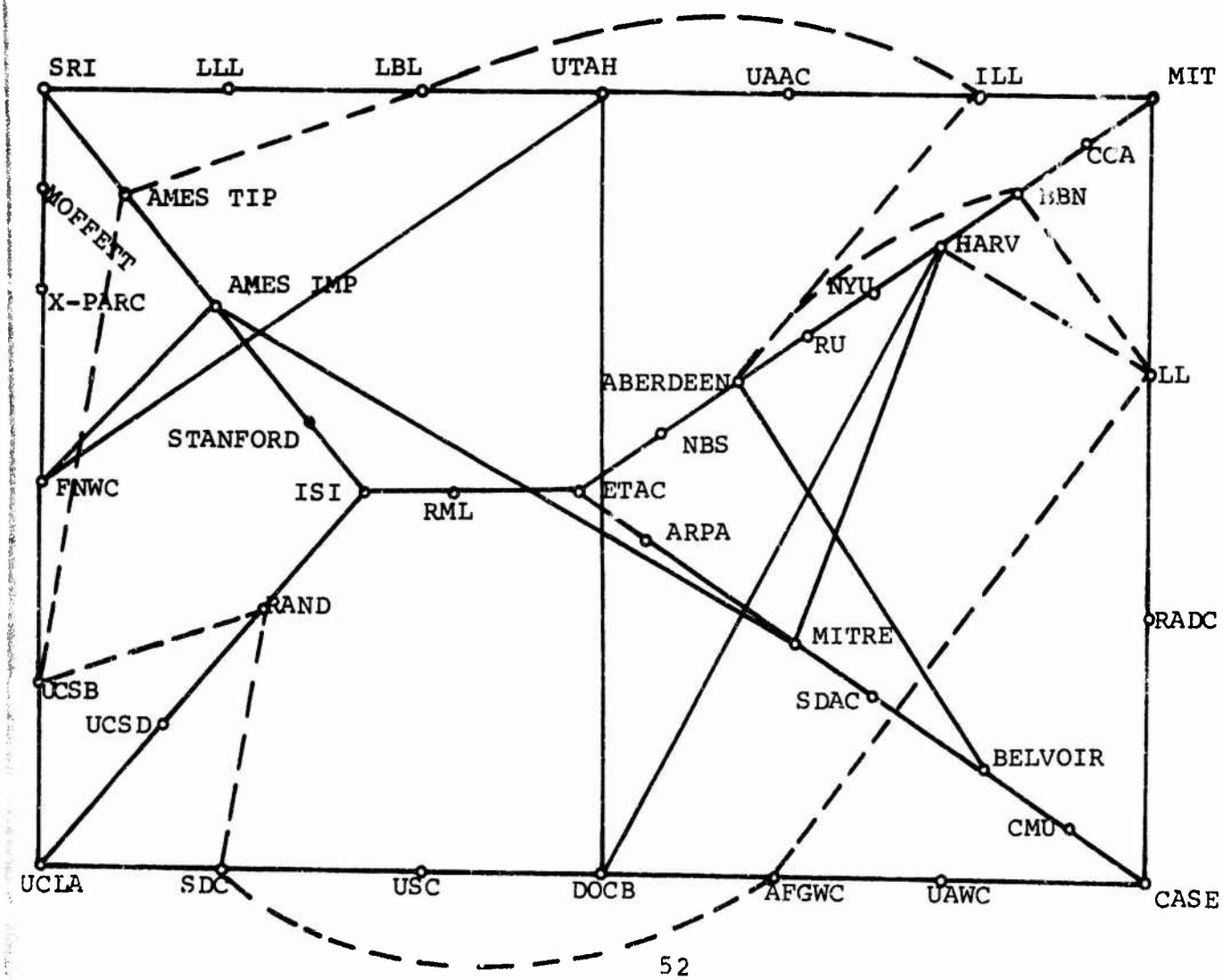
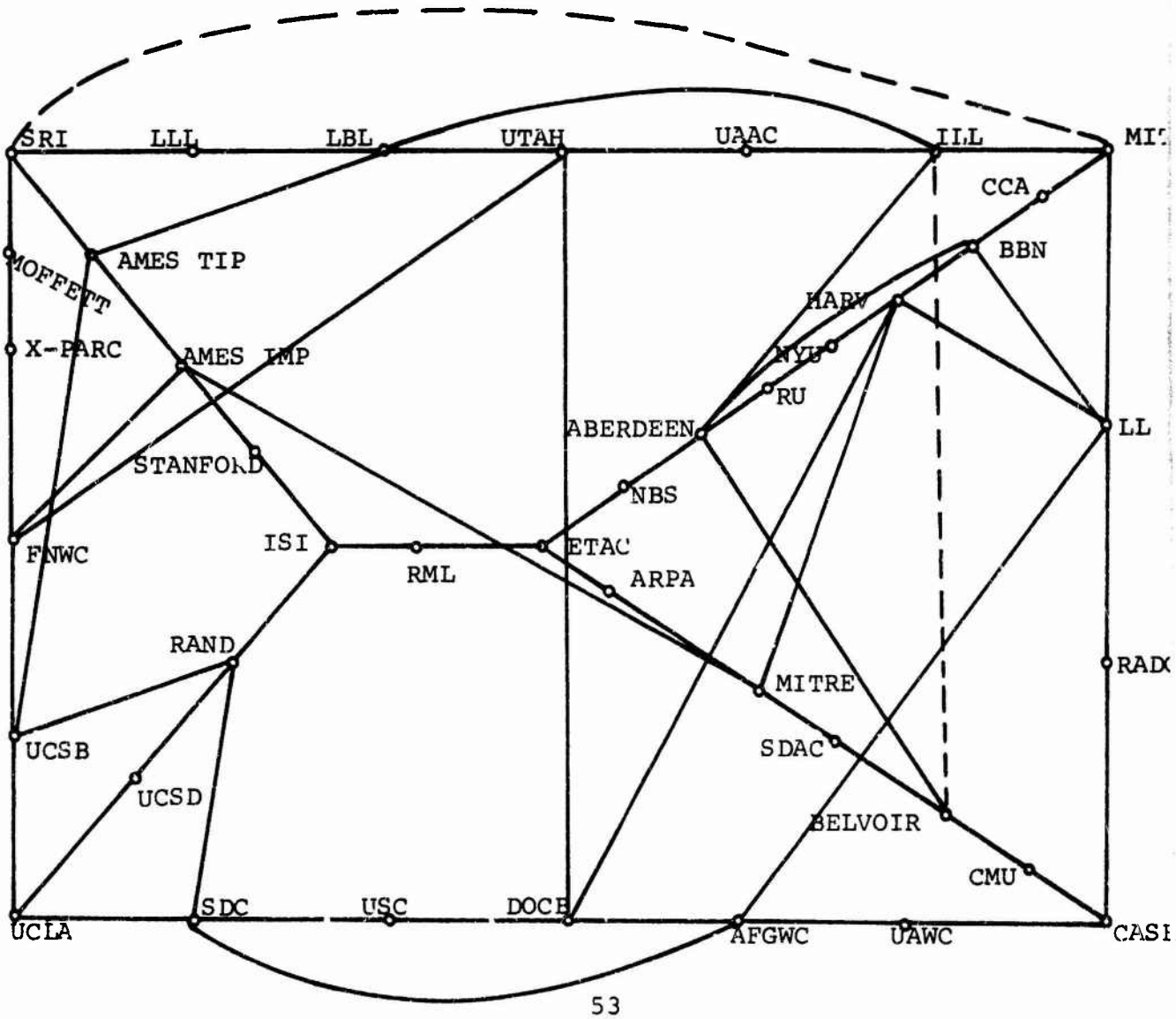


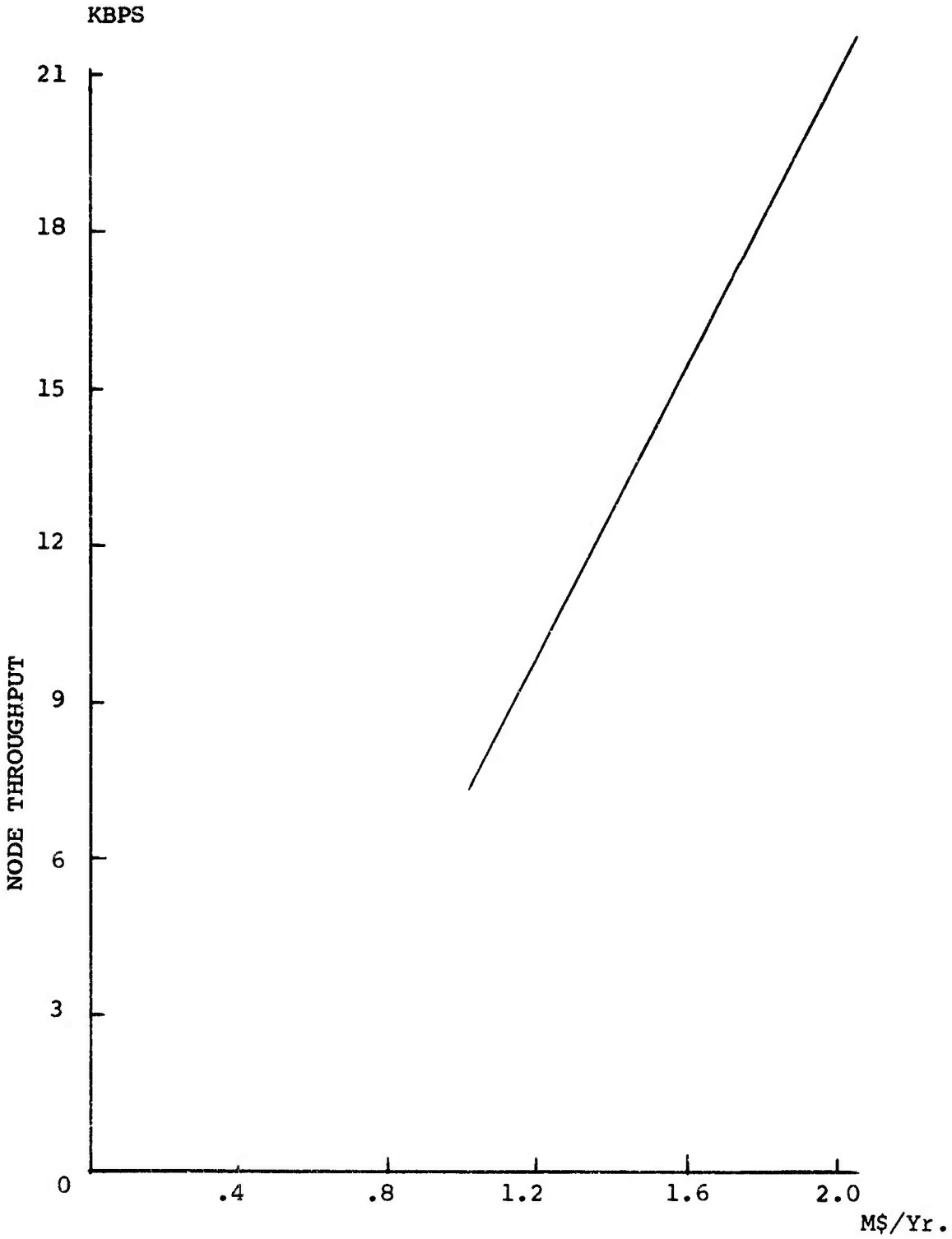
FIGURE 2.7

LINE AND MODEM COSTS: 2.053 Million Dollars

THROUGHPUT: 21.7 KBPS/Node
3.0 M-PKTS/Day/Node
120 M-PKTS/Day

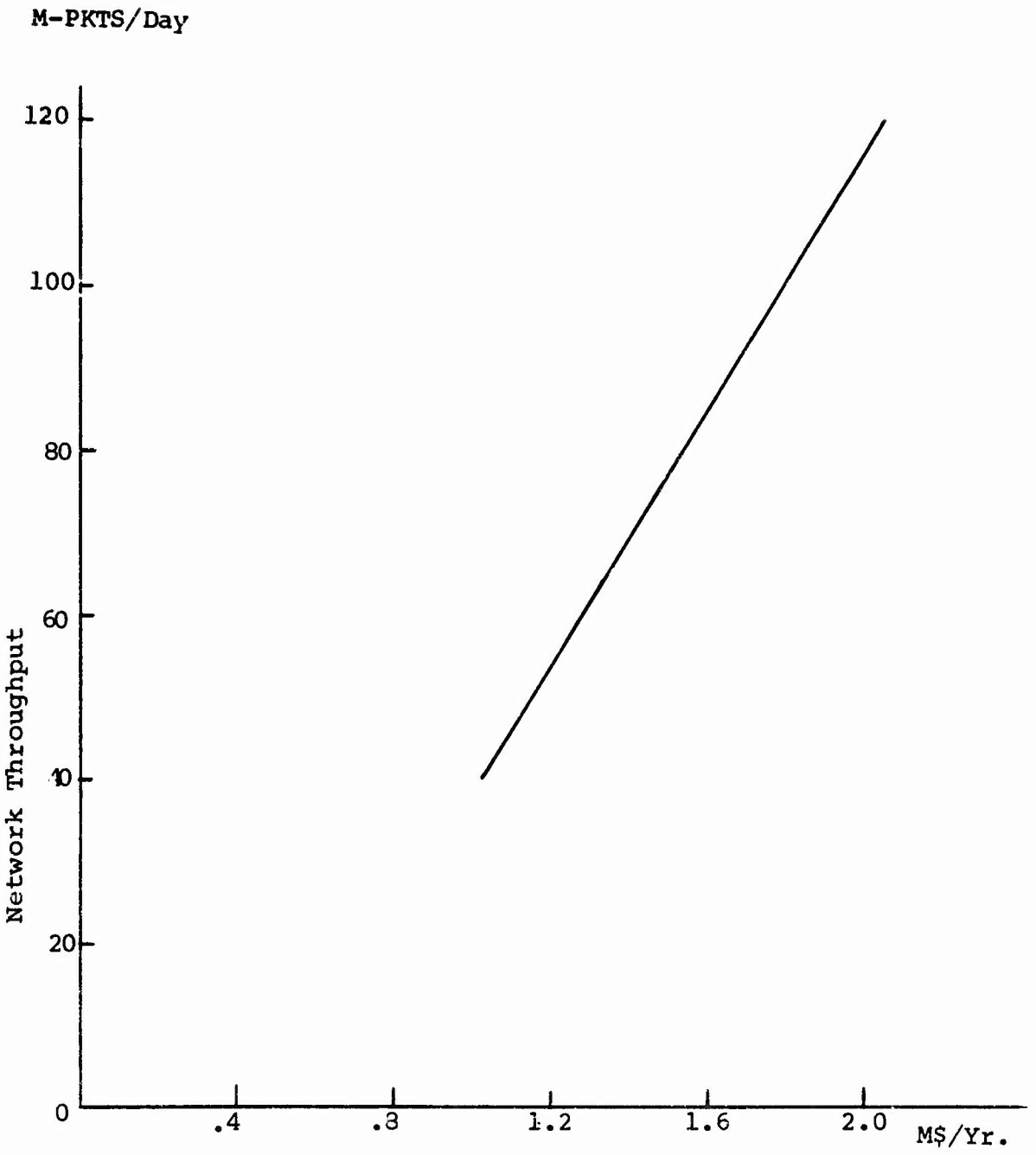
LINES ADDED: SRI-MIT
ILL-BELVOIR





NODE THROUGHPUT VS. COMMUNICATIONS COSTS

FIGURE 2.8(a)



NETWORK THROUGHPUT VS. COMMUNICATION COSTS

Figure 2.8(b)

4. TRAFFIC SENSITIVITY STUDY

4.1 Introduction

Any network design depends on the forecast of the traffic distribution. If this forecast is inaccurate, it can be expected that inefficiencies in performance will occur. The question is how sensitive the network performance is to the inaccuracy of the forecast. To investigate this question, NAC previously carried out two experiments. (Details are given in the Final Technical Report to Contract DAHC 15-70-C-0120.) Simply stated, the traffic requirement between each IMP pair is randomly generated. With the relative traffic level between different IMP pairs fixed, the maximum throughput is determined. A large number of samples were taken. The result shows that more than 75% of the random cases have average throughputs within 17% of the throughput for uniform traffic requirements. In other words, the degradation in performance caused by variations in traffic requirements is not great. In a second experiment, highly nonuniform traffic was assumed and similar conclusions were derived.

Since there were initially no accurate measurements or forecasts available for ARPANET node-node traffic requirements, some sort of assumptions were necessary. With the above experiments as a justification, NAC has been desigining and updating

ARPANET by assuming uniform traffic requirements. Even though internode traffic requirements are still not available, the packet rates out of each IMP or TIP and the packet rate on each link is now being measured. In Section 4.2, traffic requirements are derived by interpreting the NCC's December host throughput summary and line throughput summary. (The summaries are shown in Tables 2.4 and 2.5.) Based on the relative traffic level of the derived requirements, throughput is determined for the operational ARPANET of December 1972. (The network is shown in Figure 2.9 .) The throughput for uniform traffic distribution on the same network is also obtained. The two throughputs are compared.

The current throughput summaries measured by NCC do not reflect the eventual traffic pattern when the ILLIAC IV becomes fully operationa In Section 4.3, a sensitivity study is performed by assuming that 40% of the traffic is related to the ILLIAC IV and the two AMES nodes.

4.2 Current ARPANET Traffic Pattern

A. Nonuniform traffic pattern assumptions

The following are descriptions of the assumptions used in deriving the traffic requirements from the NCC's measurements.

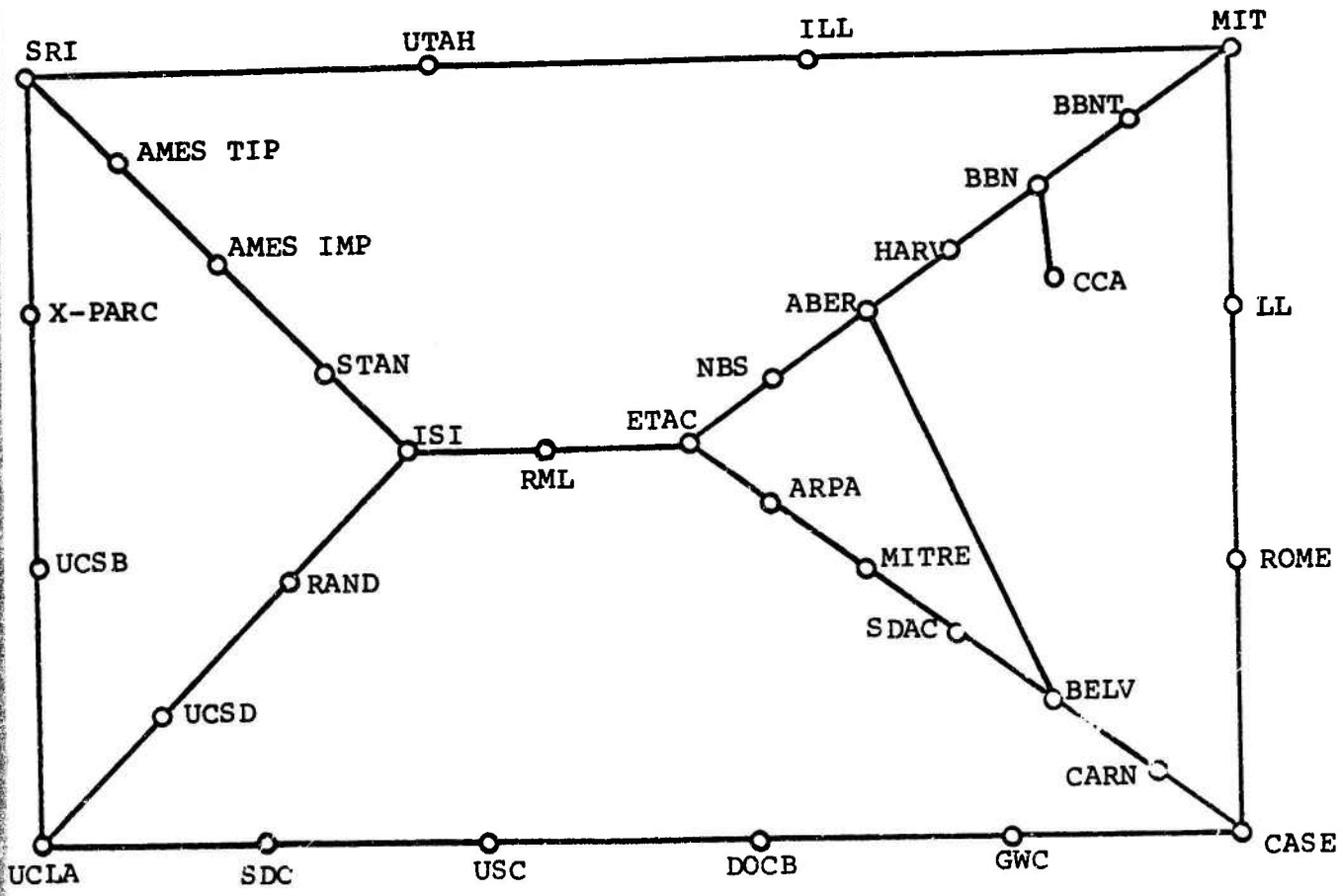

SRI
UTAH
ILL
MIT
AMES TIP
BBNT
BBN
AMES IMP
HARV
CCA
X-PARC
ABER
LL
STAN
NBS
ISI
ETAC
RML
ARPA
UCSB
MITRE
ROME
RAND
SDAC
BELV
UCSD
CARN
CASE
UCLA
SDC
USC
DOCB
GWC

FIGURE 2.9

(1) 106 Kilopackets/Day between AMES IMP and AMES TIP

Examining the measured packet rates on the links, one can see that there is a particulary high rate on the link between the AMES' TIP and AMES' IMP. One can reasonably conclude that the traffic requirement between the two is very high. Without additional information available, any reasonable assumption seems to be a good assumption. With this philosophy it is assumed that the number of packets originating from the AMES TIP and passing through the AMES IMP is the same as the number of packets originating from the AMES IMP and passing through the AMES TIP. It is further assumed that the traffic between the two is almost symmetric. With these assumptions, the traffic between the two is approximately 106 Kilopackets/Day.

(2) The traffic from one node to each of the other nodes is divided in proportion to each node's outgoing traffic.

Let TR(I,J) be the derived packet rate from node I to node J; TR(I) be the total measured packet rate generated from node I, then

$$TR(I,J) = TR(I)\ TR(J)/(\sum_{k=1}^{N} TR(k) - TR(I))$$

where N is the total number of IMPs and TIPs in the network. (The Exception to this formula is the traffic between the AMES IMP and the AMES TIP. In using the above formula, TR(I) for

the AMES IMP is the AMES IMP total traffic less the traffic from AMES IMP to AMES TIP. The same modification applies for the AMES TIP.)

The above two assumptions enable us to derive relative traffic requirements between all node pairs.

B. Results and Conclusions

- The throughput under the nonuniform traffic requirements is 9.82 KBPS/node on the average, or 44.7 million packets/day; the throughput under the uniform traffic assumption is 10.0 KBPS/node or 44.5 million packets/day. This result further strengthens our previous conclusion that the variation in traffic distribution does not in general have great effects on performance.
- If assumption (1) is removed, the throughput is found to be 6.86 KBPS or 31.2 million packets/day. The heavy traffic generated by the two AMES nodes (over 25% of the total network traffic), saturates links SRI-AMES IMP and Stanford-ISI while other links are still under-utilized. Even under this case, the throughput is only 30% below the throughput under uniform traffic requirements. This result further points out that if assumption (1) is not nearly true, the bottleneck of the network will be caused by the two AMES nodes. Therefore, the first link to be added to the network for expanding throughput should

be the one connecting AMES TIP to a node in either Boston area or Washington, D.C. area.

- Because of the closeness between network performance in both the uniform and nonuniform cases, the study presented in Section 3 for economical upgrading of the ARPANET is valid even though the actual traffic pattern deviates from the uniform one assumed.

4.3 Network Traffic Pattern with Anticipated ILLIAC IV Traffic

A. Nonuniform traffic pattern assumptions

It is assumed that

10% of the traffic is uniformly from the AMES IMP to all the non-AMES nodes;

10% of the traffic is uniformly from all the non-AMES nodes to the AMES IMP;

10% of the traffic is uniformly from the AMES TIP to all the non-AMES nodes;

10% of the traffic is uniformly from all the non-AMES nodes to the AMES TIP;

and 60% of the traffic is distributed equally between all the non-AMES nodes.

The traffic between the two AMES nodes is assumed to flow only on 230 KBPS line linking the two nodes and have no effect on the rest of the traffic in the network.

B. Results and Conclusion

(1) Applying the nonuniform traffic assumption stated above to the network shown in Figure 2.1(c), the throughput is found to be 7.3 KBPS/node or 46 M-Packets per day. This is more than 10% higher than the throughput obtained by the uniform traffic assumption. It should be noted however that this is possible only because the capacity of the AMES TIP-AMES IMP link is 230 KBPS. Otherwise, there would be a bottleneck around the AMES nodes and the throughput would have dropped by one-third.

(2) The ARPANET was not originally designed for handling high traffic volumes for AMES nodes. The logical question is then: what is the maximum throughput the network can handle by allowing local topological modifications to adapt the network for the nonuniform traffic pattern (but without any additional cross country lines)? Figure 2.10 shows such a network. The estimated lines and modem cost is about 1.167 million dollars per year--about 3% higher. The throughput is 7.93 KBPS/node, or 50 M-Packets/day--about 9% higher. Hence, it can be concluded that the current ARPANET can easily handle the possible high traffic volume generated from the AMES nodes (including the ILLIAC IV), if the traffic volume does not exceed the projected network throughput. However, if the traffic volume is to be higher, a cross country line becomes necessary.

(3) AMES related traffic in (1) and (2) is 40% of the total network traffic. This is a hypothetical number and the actual AMES related traffic may be quite different. In Figure 2.11, the curve shows the total network throughput as a function of the percentage of the AMES related traffic. The throughputs are obtained by using the network in Figure 2.10.

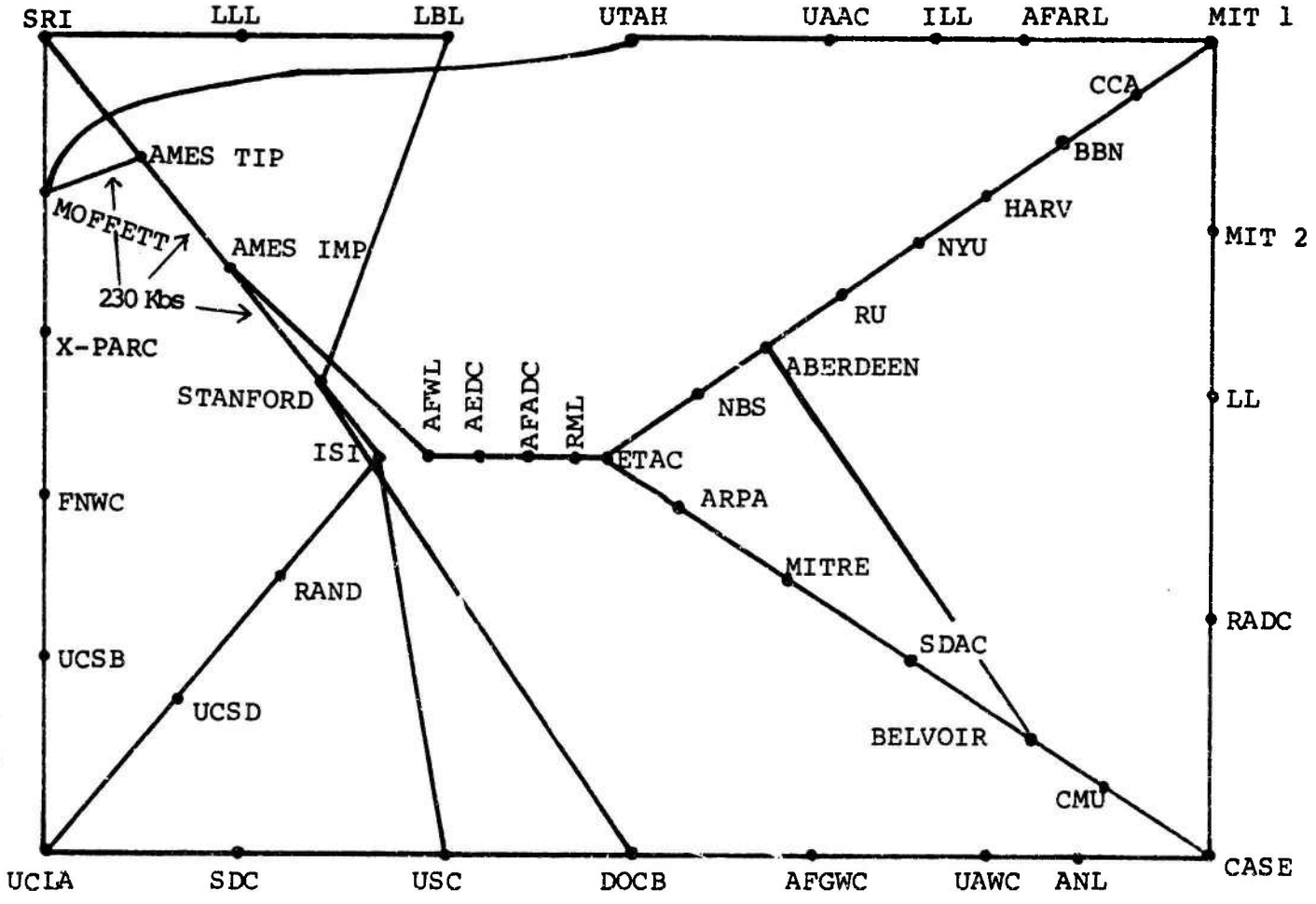

SRI
LLL
LBL
UTAH
UAAC
ILL
AFARL
MIT 1
CCA
BBN
HARV
NYU
RU
ABERDEEN
NBS
ETAC
ARPA
MITRE
SDAC
BELVOIR
CMU
CASE
RADC
LL
MIT 2
AMES TIP
MOFFETT
AMES IMP
230 Kbs
X-PARC
STANFORD
AFWL
AEDC
AFADC
RML
ISI
FNWC
RAND
UCSB
UCSD
UCLA
SDC
USC
DOCB
AFGWC
UAWC
ANL

FIGURE 2.10

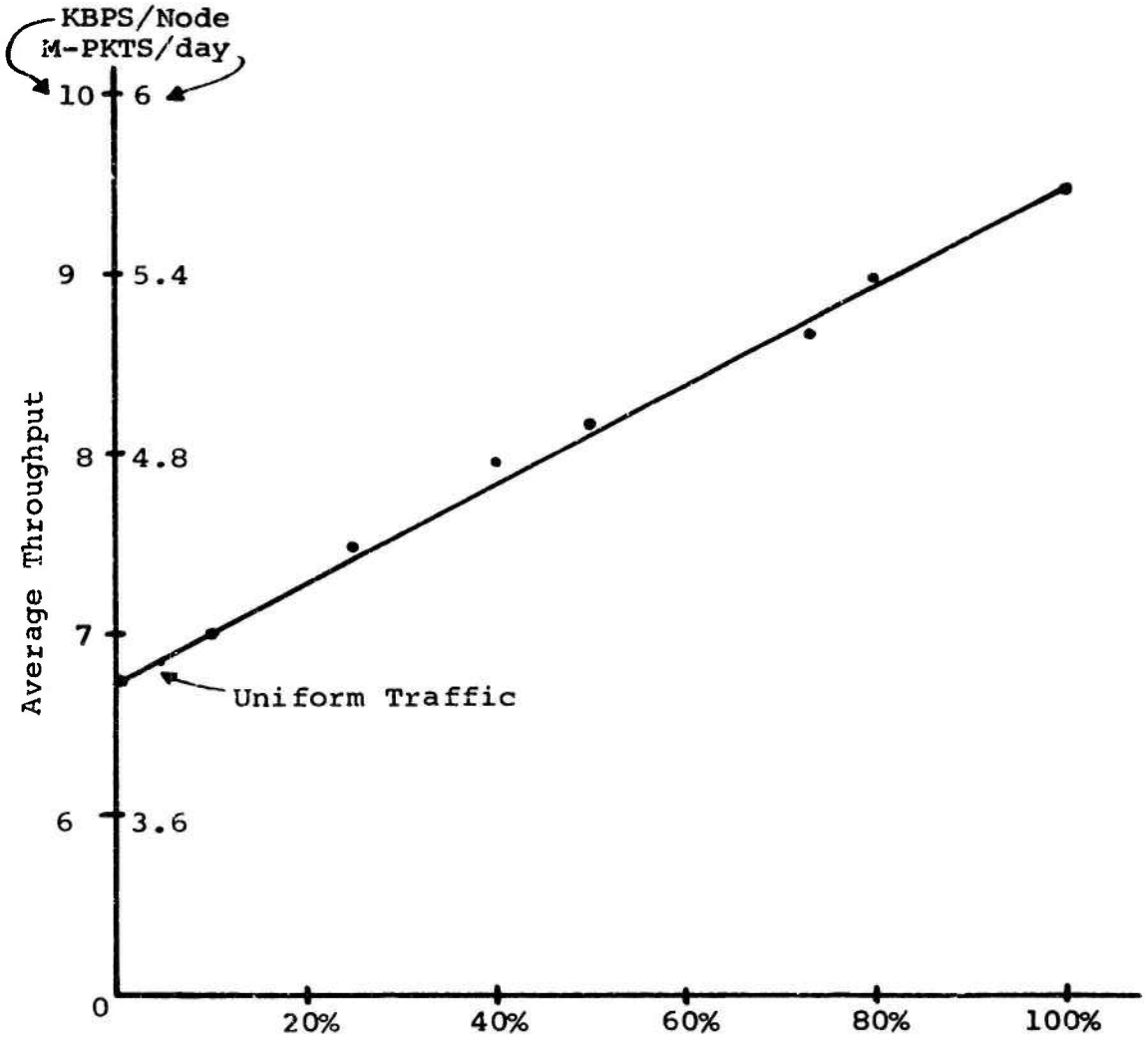

KBPS/Node
M-PKTS/day
10
6
9
5.4
8
4.8
7
6
3.6
0
Average Throughput
Uniform Traffic
20%
40%
60%
80%
100%
AMES related traffic / Total network traffic x 100

FIGURE 2.11

Chapter 2–Appendix A

TABLE A-I

HOST THROUGHPUT SUMMARY
(PACKETS)
(DECEMBER 1972)

		INTER-NODE	INTRA-NODE	TOTAL	AVG. DAILY INTERNODE	DAYS
UCLA	HOST 1	181141	57162	238303		
UCLA	HOST 2	668578	44616	713194		
		849719	101778	951497	33989	25
SRI	HOST 1	1396436	17556	1413992		
SRI	HOST 2	31365	3836	35201		
		1427801	21392	1449193	57112	25
UCSB	HOST 1	545165	173928	719093	21807	25
UTAH	HOST 1	536453	35616	572069	21458	25
BBN	HOST 2	2424250	373106	2797356		
BBN	HOST 3	108384	29878	138262		
BBN	HOST 4	153239	37	153276		
		2685873	403021	3088894	107435	25
MIT	HOST 1	48752	13988	62740		
MIT	HOST 2	482445	224148	706593		
MIT	HOST 3	401599	373678	775277		
MIT	HOST 4	201840	439670	641510		
		1134636	1051484	2186120	45385	25
RAND	HOST 1	264812	86	264898	10592	25
SDC	HOST 1	27964	4416	32380	1119	25
HARV	HOST 1	299236	397614	696850		
HARV	HOST 2	23247	315178	338425		
		322483	712792	1035275	12899	25
LINC	HOST 1	10158	1104	11262		
LINC	HOST 2	28589	67907	96496		
LINC	HOST 3	578	4512	5090		
		39325	73523	112848	1573	25
STAN	HOST 1	1015835	13196	1029031	40633	25
ILL	HOST 1	386634	280	386914	15465	25
CASE	HOST 1	221014	28570	249584	8341	25

TABLE A-I (Continued)

CARN	HOST 1	317017	217871	534888		
CARN	HOST 2	265575	193101	458676		
		582592	410972	993564	23304	25
AMES2	HOST 1	2811857	28592	2840449	112474	25
AMES1	HOST 1	57368	14650	72018		
AMES1	HOST 3	4384317	18336	4402653		
		4441685	32986	4474671	177667	25
MITRE	HOST 3	1091812	51	1091863	43672	25
ROME	HOST 3	487923	25	487948	19517	25
NBS	HOST 1	738	24	762		
NBS	HOST 3	921496	92	921588		
		922234	116	922350	36889	25
ETAC	HOST 3	878406	416	878822	35136	25
TINK HAD NO TRAFFIC						25
ISI	HOST 2	4176170	46977	4223147	167047	25
USC	HOST 1	13611	151120	164731		
USC	HOST 3	922350	146673	1069023		
		935961	297793	1233754	37438	25
GWC	HOST 3	15327	76	15403	613	25
NOAA	HOST 3	58916	29885	88801	2357	25
SAAC	HOST 3	209321	538	209859	8373	25
BELV HAD NO TRAFFIC						25
ARPA	HOST 1	0	180178	180178		
ARPA	HOST 3	155445	177439	332884		
		155445	357617	513062	6218	25
ABER HAD NO TRAFFIC						25
BBN T	HOST 3	640257	3939	644196	25610	25
CCA	HOST 1	383002	318035	701037		
CCA	HOST 3	875814	148241	1024055		
		1258816	466276	1725092	50353	25
XEROX HAD NO TRAFFIC						25
UCSD	HOST 1	328825	4186	333011	13153	25

TABLE A-I (Concluded)

HAW I HOST 3	36487	38789	75276	3317	11

TOTAL	28489750	4339316
DAILY AVERAGE	1139590	173573
AVERAGE PER NODE-DAY	34079	5191
PACKETS/MESSAGES (INTERNODE)	1.06	

TABLE A-II

LINE THROUGHPUT SUMMARY
(PACKETS)
(DECEMBER 1972)

LINE NO.	SITE #1	SITE #2	AVERAGE DAILY 1 - 2	AVERAGE DAILY 2 - 1	AVERAGE DAILY TOTAL	LINE USE BUSIEST DIRECTION	DAYS
1	ETAC	TINK	234364	238315	472679	7.26%	25
2	MIT	BBN T	224428	174479	398907	6.84%	25
3	MIT	ILL	269921	270664	540585	8.24%	25
4	USC	NOAA	45708	43874	89581	1.39%	25
5	UCLA	SDC	93400	94174	187574	2.87%	25
6	UCLA	UCSD	129795	128053	257848	3.95%	25
7	UCLA	UCSB	104968	96942	201911	3.20%	25
8	SRI	XEROX	106902	114425	221327	3.49%	25
9	BELV	ABER	79988	66998	146987	2.44%	22
10	SRI	UTAH	325362	323272	648634	9.91%	25
11	SRI	AMES1	280274	278694	558968	8.54%	25
12	STAN	ISI	276966	275693	552659	8.44%	25
13	BBN	HARV	114261	116356	230617	3.54%	25
14	MIT	LINC	139392	134056	273448	4.25%	25
15	UTAH	ILL	316268	306703	622971	9.63%	25
16	LINC	ROME	131597	125990	257587	4.01%	25
17	CASE	CARN	74895	70493	145388	2.28%	25
18	CARN	BELV	60944	54958	115902	1.86%	19
19	HARV	ABER	97819	79885	177704	2.98%	25
20	STAN	AMES2	278144	281708	559852	8.58%	25
21	MITRE	ARPA	106591	105699	212491	3.25%	23
22	CASE	ROME	90813	96437	187250	2.94%	25
23	NBS	ETAC	110756	111578	222335	3.40%	25
24	AMES2	AMES1	490831	494211	985042	15.05%	25
25	TINK	ISI	237610	238740	476351	7.27%	25
26	SDC	USC	90734	91348	182082	2.78%	25
27	CASE	GWC	68039	66686	134725	2.07%	17
28	SAAC	BELV	85480	86389	171869	2.63%	22
29	MITRE	SAAC	80128	80934	161062	2.47%	25
30	BBN	BBN T	220819	160104	380924	6.73%	23
31	BBN	CCA	103945	104048	207993	3.17%	24
32	GWC	NOAA	63080	61948	125028	1.92%	18
33	ETAC	ARPA	113458	114122	227581	3.48%	23
34	NBS	ABER	85383	53590	138973	2.60%	14
35	UCSB	XEROX	117183	109360	226543	3.57%	24
36	RAND	ISI	123840	125646	249485	3.83%	25
38	RAND	UCSD	118517	122732	241249	3.74%	25
39	AMES2	HAW T	8761	8548	17309	.27%	11

AVERAGE (ONE-DIRECTIONAL) LINE THROUGHPUT = 153248

AVERAGE LINE UTILIZATION = 4.67%

3. PROPERTIES OF LARGE NETWORKS--PART I

1. INTRODUCTION

A long range objective of the present study effort is to investigate the feasibility and merits of packet switching for widespread Defense Department application. NAC's past studies of packet switching have demonstrated the viability of the ARPANET approach for systems with as many as 100 to 200 nodes. Defense Department communications and computational requirements are global in scope and immensely complex. Prior to the present study effort, there has been little if any systematic study of packet switching for systems of the size that could make a substantial impact on Defense Department operations.

Analysis and design of large scale networks requires techniques substantially different from the ones used for smaller networks. Furthermore, the adaptive routing techniques currently implemented in the ARPANET cannot be directly utilized in a very large network because of excessive IMP processing time, memory requirements and traffic overhead. Consequently, before a general attack on the large network analysis and design problem is sensible, it is desirable to show the existence of workable networks that are able to meet time delay, traffic and reliability requirements.

Demonstrating the existence of such networks and an analysis of their cost-reliability throughput characteristics can establish the validity of the overall approach.

In this chapter, we present the results of the first large network packet switched design effort ever attempted. The network was chosen to contain 1,000 nodes since this is an order of magnitude greater than any other system design ever attempted. The primary approach selected utilizes a hierarchical network implementation, in which various subnetworks are designed and operated using traditional techniques. Because of the complexity of the design optimization (which involves the determination of 100 node partitions and the solution of 111 network subproblems!), only feasible, low cost solutions are generated. The analysis is performed using a decomposition approach. Cost, throughput, delay and reliability are evaluated for each subnetwork; the overall network performance is then obtained by properly combining the partial results.

Two simple, non-hierarchical 1000 node structures are also considered and their cost and performance compared to the hierarchical case. It is shown that the hierarchical structure exhibits lower cost, and offers more flexibility in the design and easier control of network reliability.

The comparison of the 1000 node hierarchical results with those already available for networks of sizes up to 200 nodes shows that cost, throughput and reliability of the 1000 node network follow the trends identified in smaller size networks. In particular, it is shown that reliability requirements become more critical to satisfy, where network size increases; in a 1000 node network, for example, satisfactory reliability can be achieved in general only with 3-connected topologies.

2. SYSTEM PARAMETERS AND CHARACTERISTICS

The following is a list of the factors that influence the network design.

1) The system contains 1000 Message Processors located in the largest cities of the Continental United States. The number of Message Processors in each town is proportional to the population of the town. The map in Figure 3.1 displays the locations and the number of Processors per location.

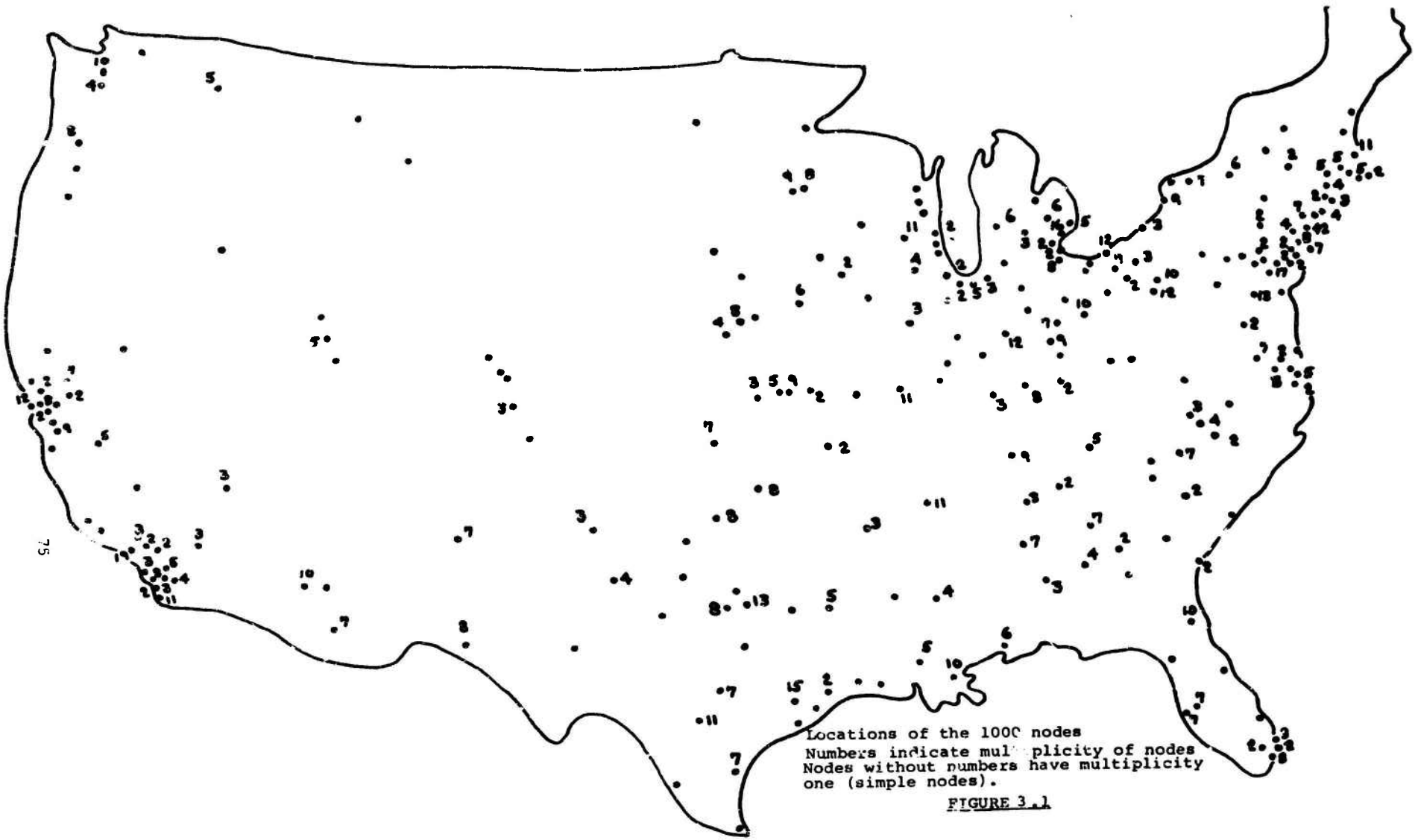
2) Required traffic between Message Processors is assumed uniform for all node pairs. Traffic levels in the range from 3 to 20 Kbs/node are considered.

3) Messages are assumed to have the same structure and formats as in the present ARPANET configuration. Message delay is evaluated for single packet messages.

4) The nominal traffic level is set at 80% of the saturation level in order to maintain within acceptable limits the queue size of packets awaiting transmission on each channel.

5) The link failure rate is assumed equal to 0.02. The node failure rate is assumed equal to 0.02 for IMP and TIP processors, and .0004 for redundant configurations (IMP or TIP plus backup, or redundant high speed modular IMP configurations).

6) The high throughput presented by a 1000 node network requires very high channel and message processor rates. Therefore,



Locations of the 1000 nodes
Numbers indicate multiplicity of nodes
Nodes without numbers have multiplicity
one (simple nodes).

FIGURE 3.1

in the design, two high speed hardware options--the 1544 Kbs data channel and the HSMIMP (High Speed Modular IMP)--have been considered in addition to the options already available. Such high rate options are presently under development but are not yet operational offerings. Hardware costs and characteristics are shown in Tables 3.1 and 3.2.

TABLE 3.1

LINE COSTS

Capacity (Kbs)	Data Set Cost/Month	Line Cost per Mile/Month
9.6	$493	$0.42
19.2	$850	$2.50
50.0	$850	$5.00
230.4	$1300	$30.00
1544.0	$2000	$75.00

All lines full duplex

TABLE 3.2

MESSAGE PROCESSOR COST

Description	Purchase Cost	Cost/Year*
DDP-316 IMP (Max throughput = 600 Kbs)	$50,000	$15,000
DDP-516 IMP (Max throughput = 800 Kbs)	$70.000	$21,000
DDP-316 TIP (Max throughput < 600 Kbs)	$100,000	$30,000
HSMIMP (Max throughput = 6,000 Kbs)	$250,000	$75,000

* Yearly cost is assumed 30% of purchase cost

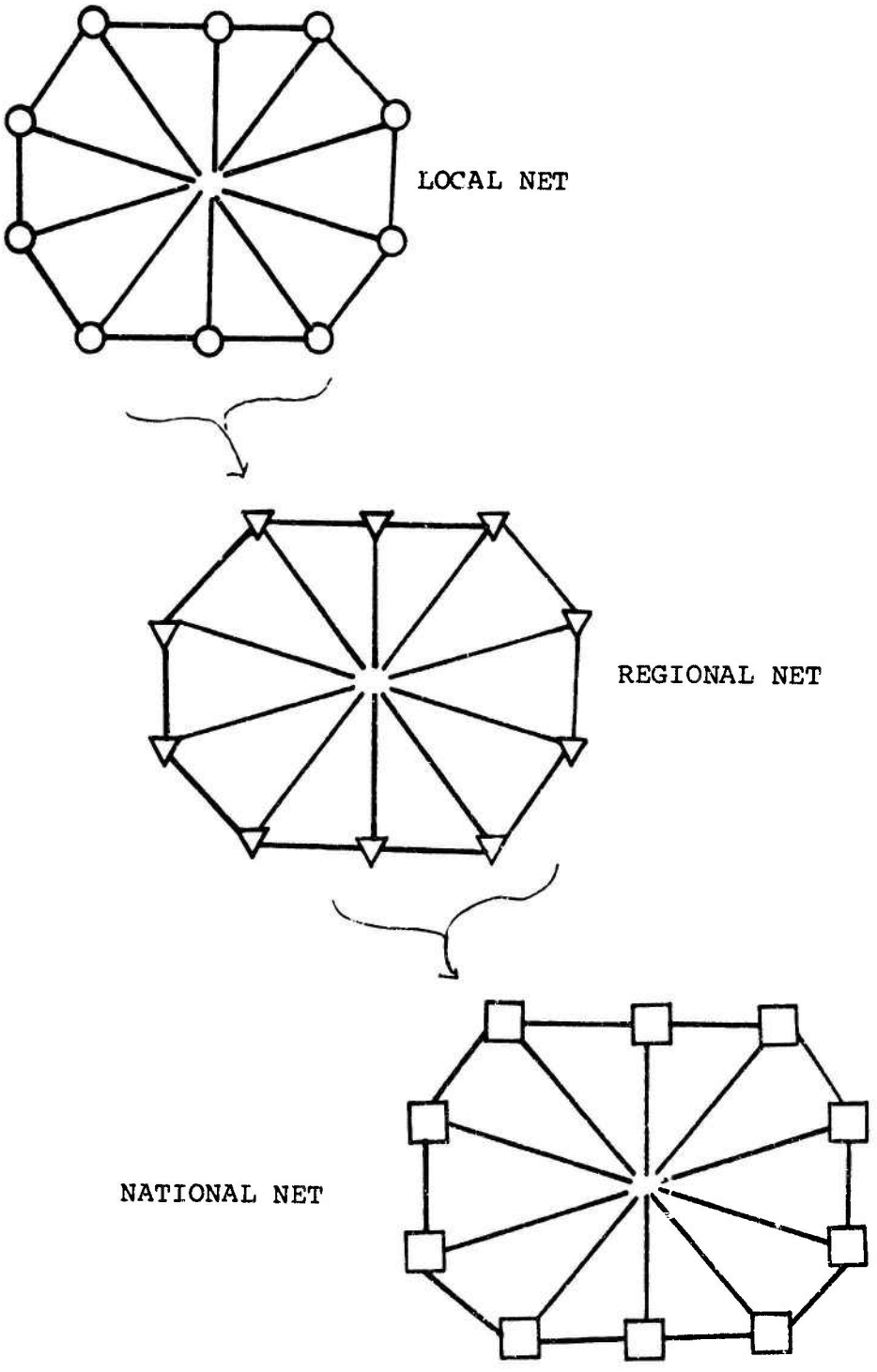
3. HIERARCHICAL NETWORK STRUCTURE

3.1 Network Topology

The determination of the optimal topology in a 1000 node hierarchical network is a very complex problem, as it requires the solution of a large number of subproblems, all connected with each other. For example, one must optimally determine: (1) the number of hierarchical levels, (2) the node partitions, (3) the topology within each partition, (4) the connections between networks in different hierarchical levels.

Due to the complexity of the optimal design, only a feasible, reasonably low cost design was considered in the first phase of the 1000 node study. A feasible design in fact is sufficient for the determination of cost, throughput, delay and reliability trends with respect to network size, and for a comparison between hierarchical and non-hierarchical structures.

The hierarchical structure here considered consists of three hierarchical levels: one 10 node national network, ten 10 node regional networks, and one hundred 10 node local networks (see Figure 3.2). Each local network is considered as one "node" of the higher level regional net, and similarly each regional net is one node of the national net. Various ways of connecting lower to higher level networks can be considered. In the cost-



A 1000 node network composed of 10 ten-node regional nets each containing 10 ten node local nets.

FIGURE 3.2

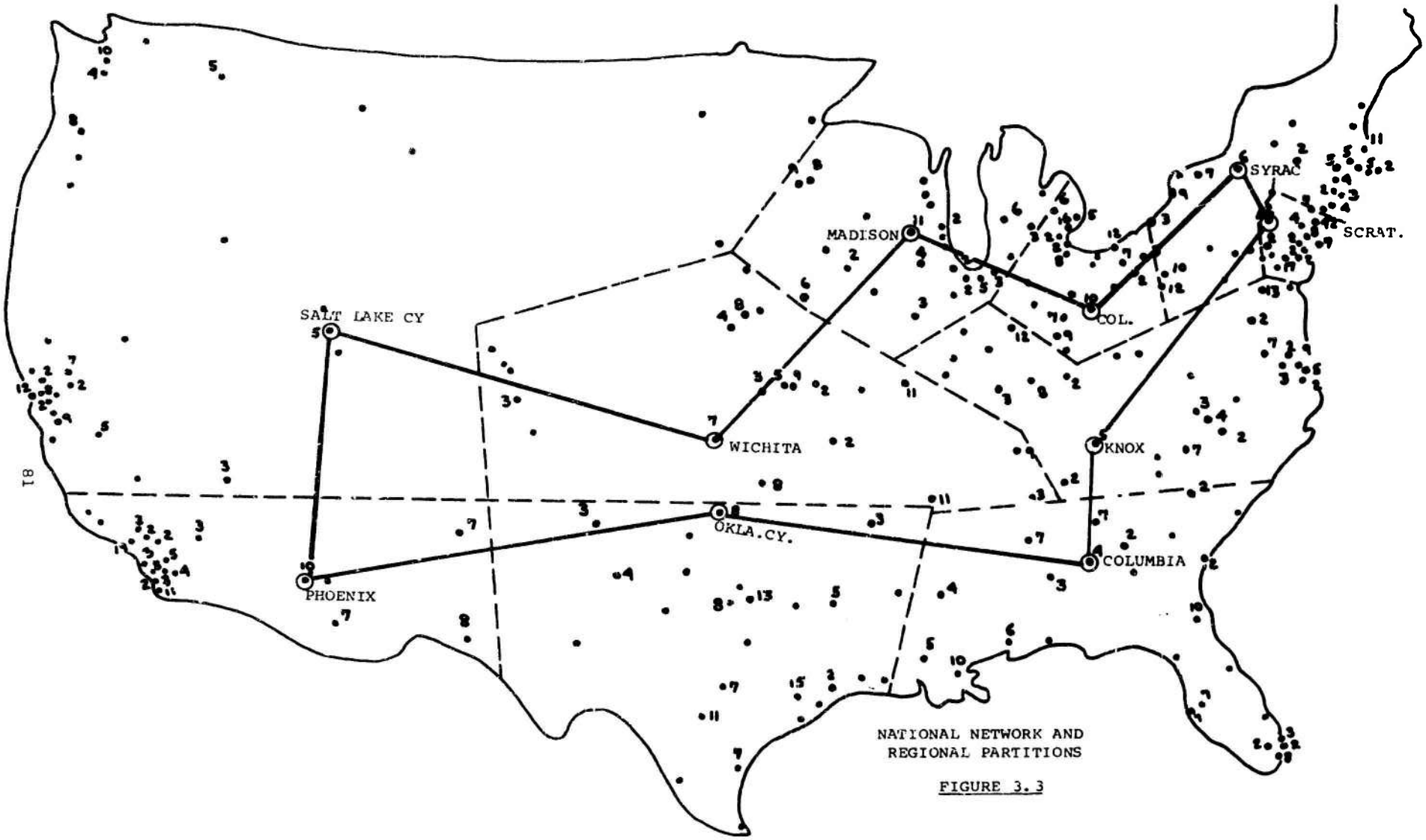
throughput study, we assume for simplicity that each subnetwork communicates with the higher level network only through one "exchange" node. In the reliability study, however, also two and three exchange node configurations are considered.

In order to achieve an acceptable reliability, national and regional networks are 3-connected, as in Figure 3.2. For the local networks, which contribute to the total communication cost by more than 50%, also less expensive configurations, which are not 3-connected, were investigated.

Figure 3.3 shows the 10 regional partitions and the outer loop of the national network topology. Figure 3.4 shows the 10 local partitions in the region that covers Texas and outer loops of regional and local topology.

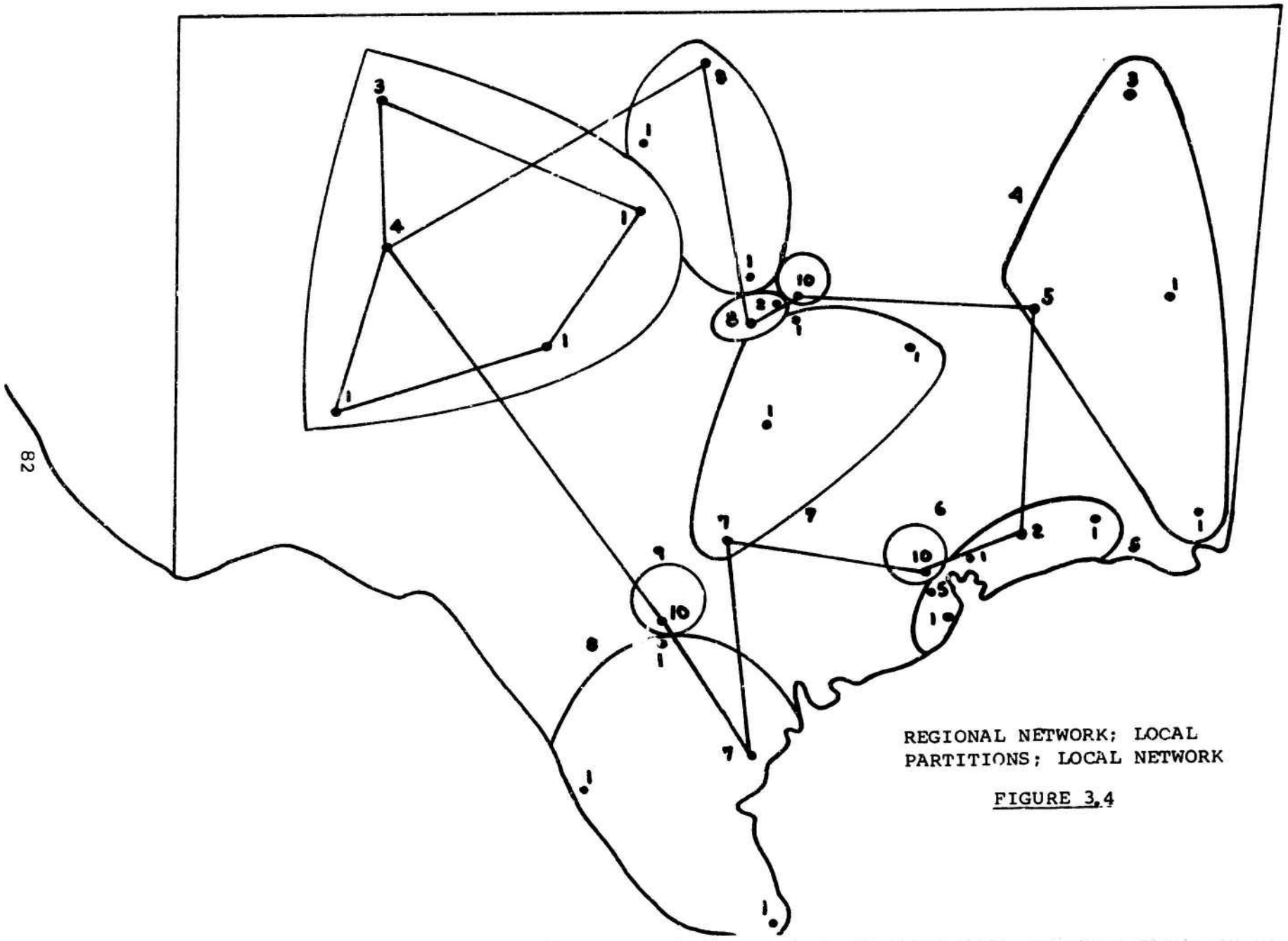
3.2 Throughput, Cost and Delay

Total cost and delay for a given throughput can be obtained by analyzing 111 subnetworks and properly combining the results. Such an extensive analysis is too cumbersome in our case since we are interested in a parametric study, using the throughput as a parameter. Therefore, in order to simplify the computation, only the national network shown in Figure 3.3 and the regional and local nets shown in Figure 3.4 were thoroughly analyzed, and the results interpreted as representative for all the other



NATIONAL NETWORK AND
REGIONAL PARTITIONS

FIGURE 3.3



REGIONAL NETWORK; LOCAL
PARTITIONS; LOCAL NETWORK

FIGURE 3.4

regional and local nets. Notice that the above approach generates imprecision in the total cost, but provides the correct answers for both delay and throughput.

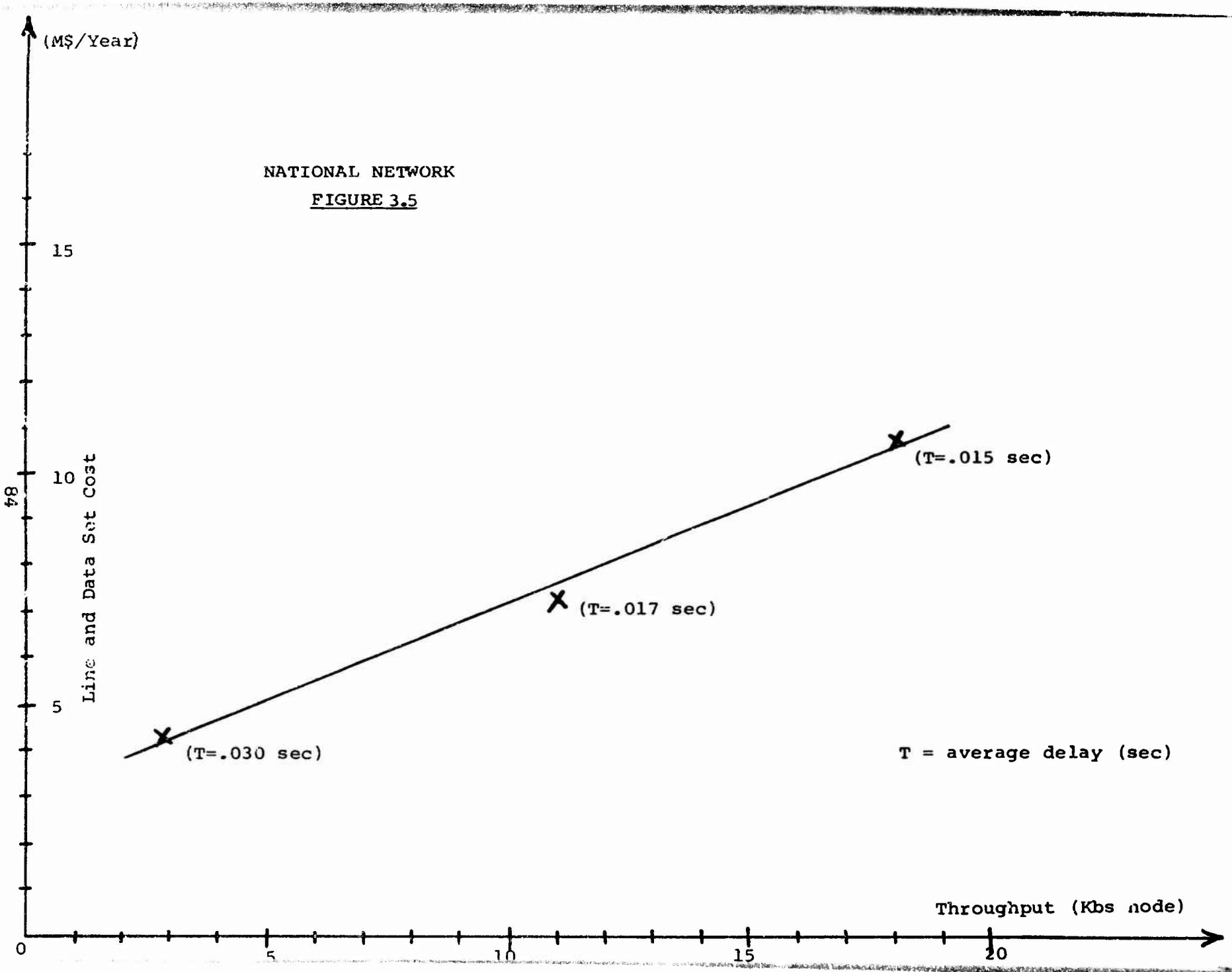
Figure 3.5 shows cost, throughput and delay of the national network for three different capacity allocations. The lowest cost configuration uses all 230.4 Kbs channel capacities. The intermediate configuration uses 1544 Kbs channels for the outer loop, and 230.4 Kbs for the cross links. The highest cost configuration uses all 1544 Kbs channels. The throughput, expressed in Kbs/node, refers to the local nodes; therefore, the throughput of each of the 10 "supernodes" in the national net is approximately 100 times higher. The cost in Figure 3.5 reflects line and data set costs. The additional message processor cost is now evaluated, assuming that eacy node has redundant processors:

(1) lower cost net: 20 x DDP-316 IMPs, cost = .3M$/year

(2) intermediate cost net: 20 x HSMIMPs, cost = 1.5 M$/year

(3) higher cost net: 20 x HSMIMPs, cost = 1.5 M$/year

Figure 3.6 shows the results for the regional net. The lowest cost solution uses mostly 50 Kbs channels; the highest cost solution includes several 230.4 Kbs and 1544 Kbs channels. The throughput refers to local nodes. Assuming that each node has redundant processors, the message processor cost is given below:


(M$/Year)
NATIONAL NETWORK
FIGURE 3.5
15
10
5
0
Line and Data Set Cost
84
(T=.015 sec)
(T=.017 sec)
(T=.030 sec)
T = average delay (sec)
Throughput (Kbs node)
5
10
15
20

(1) lower cost net: 18 x DDP-316 IMPs, cost = 270 K$/year

(2) intermediate cost net: 16 x DDP-316 IMPs, cost = 240 K$/year

2 x HSMIMPs, cost = 150 K$/year

Total cost = 390 K$/year

(3) highest cost net: 12 x DDP-316 IMPs, cost = 180 K$/year

6 x HSMIMPs, cost = 450 K$/year

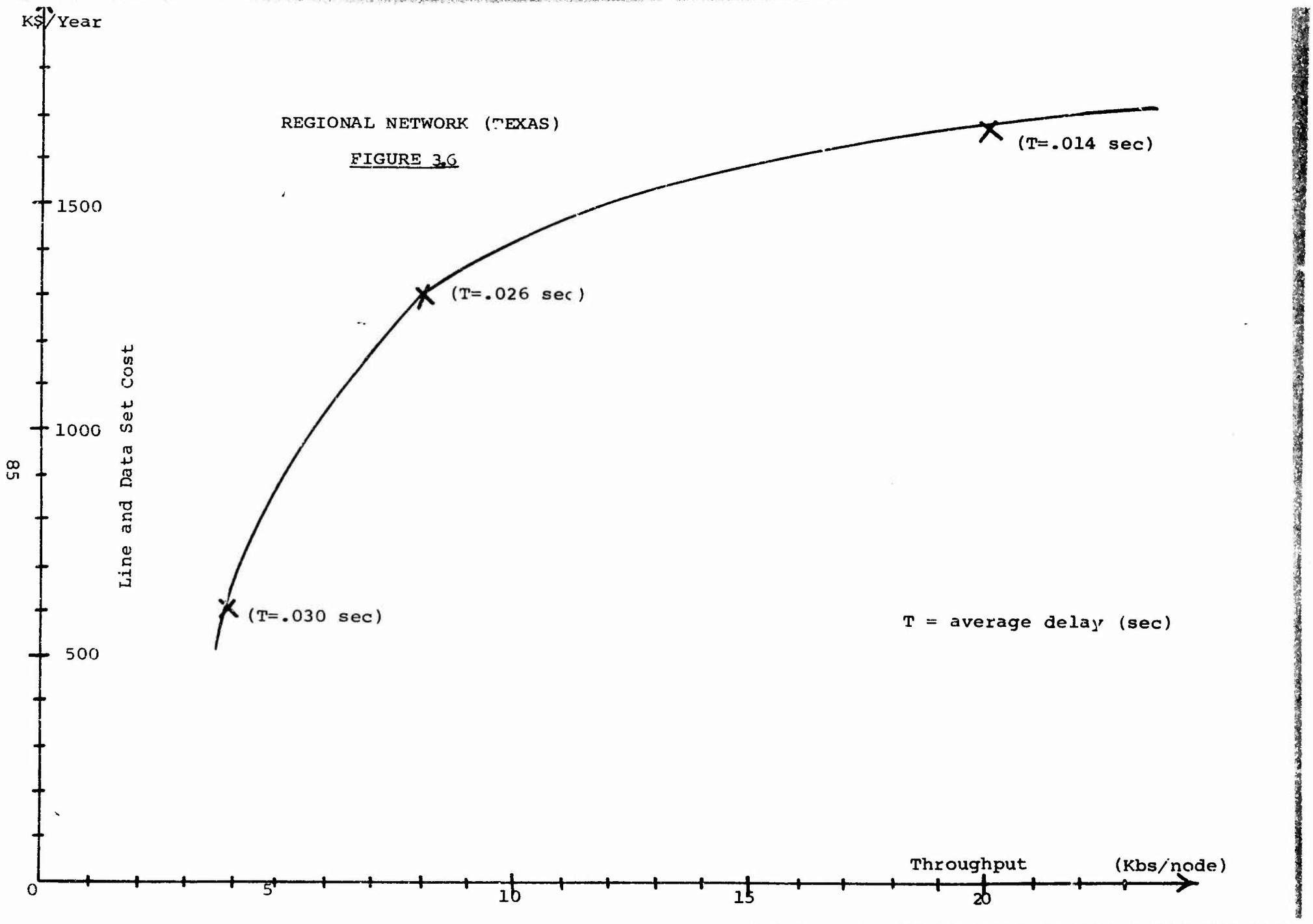
Total cost = 630 K$/year

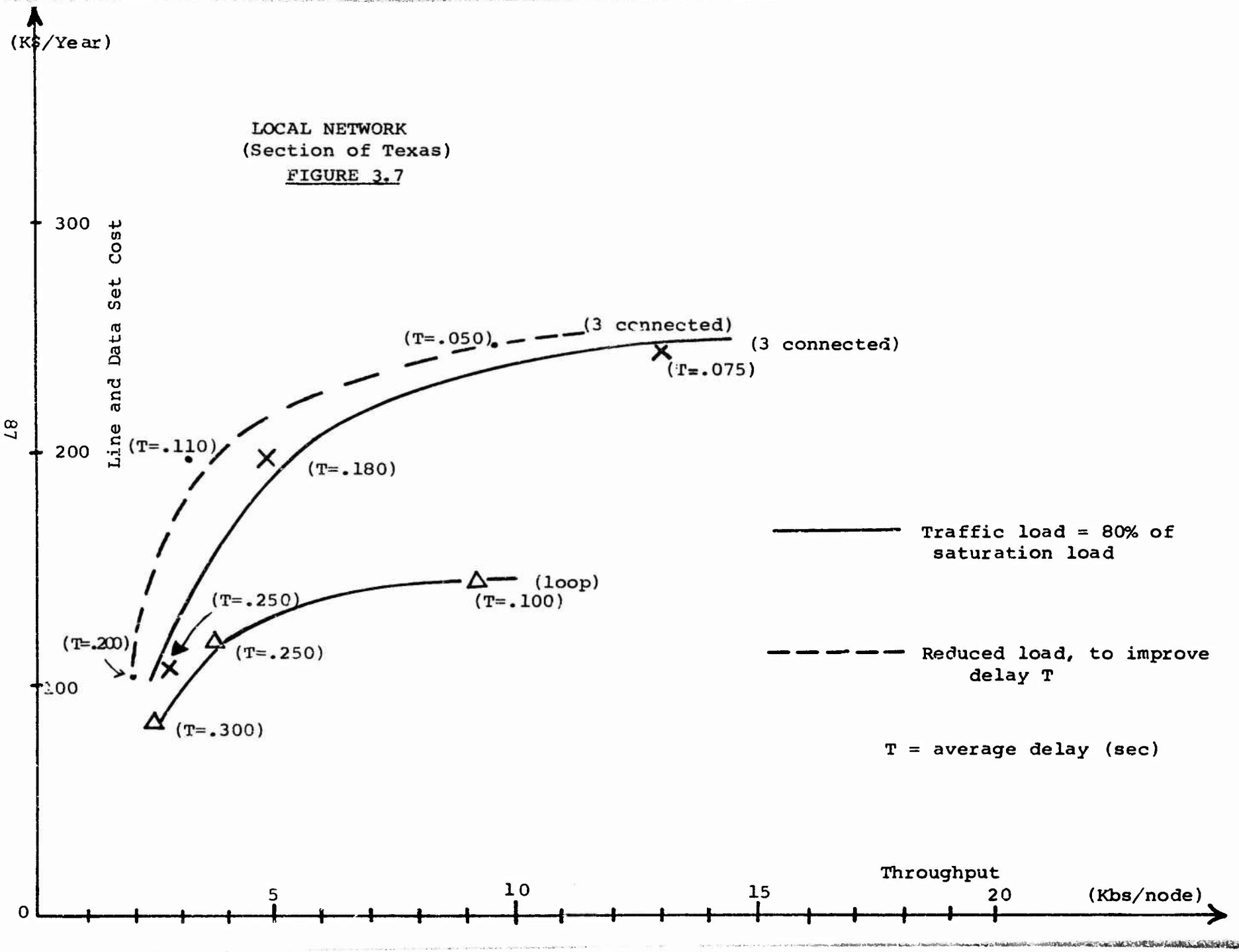
Figure 3.7 shows the results for the local net. Both 3-connected and loop configurations were analyzed. Various capacity assignments, leading to different solutions, were considered. Average delay T in the local nets is much higher than in the national and global nets, because of the extensive use of 9.6 Kbs and 19.2 Kbs channels, especially in the low cost, low throughput configurations. The delay can be reduced by reducing the traffic load, as shown in Figure 3.7. The local network does not require, in general, redundant processors; the message processor cost is given by:

(1) local network: 9 x DDP-316 IMPs, cost = 135 K$/year.

The results for the global net are obtained as follows:

(1) for each throughput level, the lowest cost national, regional and local solutions that can accommodate such a throughput are selected.


K$/Year
REGIONAL NETWORK (TEXAS)
FIGURE 3.6
1500
1000
500
0
5
10
15
20
Throughput
(Kbs/node)
Line and Data Set Cost
85
(T=.014 sec)
(T=.026 sec)
(T=.030 sec)
T = average delay (sec)


(K$/Year)
LOCAL NETWORK
(Section of Texas)
FIGURE 3.7
300
Line and Data Set Cost
87
200
100
0
(T=.050)
(3 connected)
(3 connected)
(T=.075)
(T=.110)
(T=.180)
(T=.250)
(loop)
(T=.100)
(T=.200)
(T=.250)
(T=.300)
Traffic load = 80% of saturation load
Reduced load, to improve delay T
T = average delay (sec)
5
10
15
20
Throughput
(Kbs/node)

(ii) the total cost D_t is given by:

$$D_t = D_n + 10\ D_r + 100 \times D_\ell$$

where D_n = national net cost

D_r = regional net cost

D_ℓ = local net cost

(iii) the total average delay T_t suffered by a packet traveling from source to destination is typically given by:

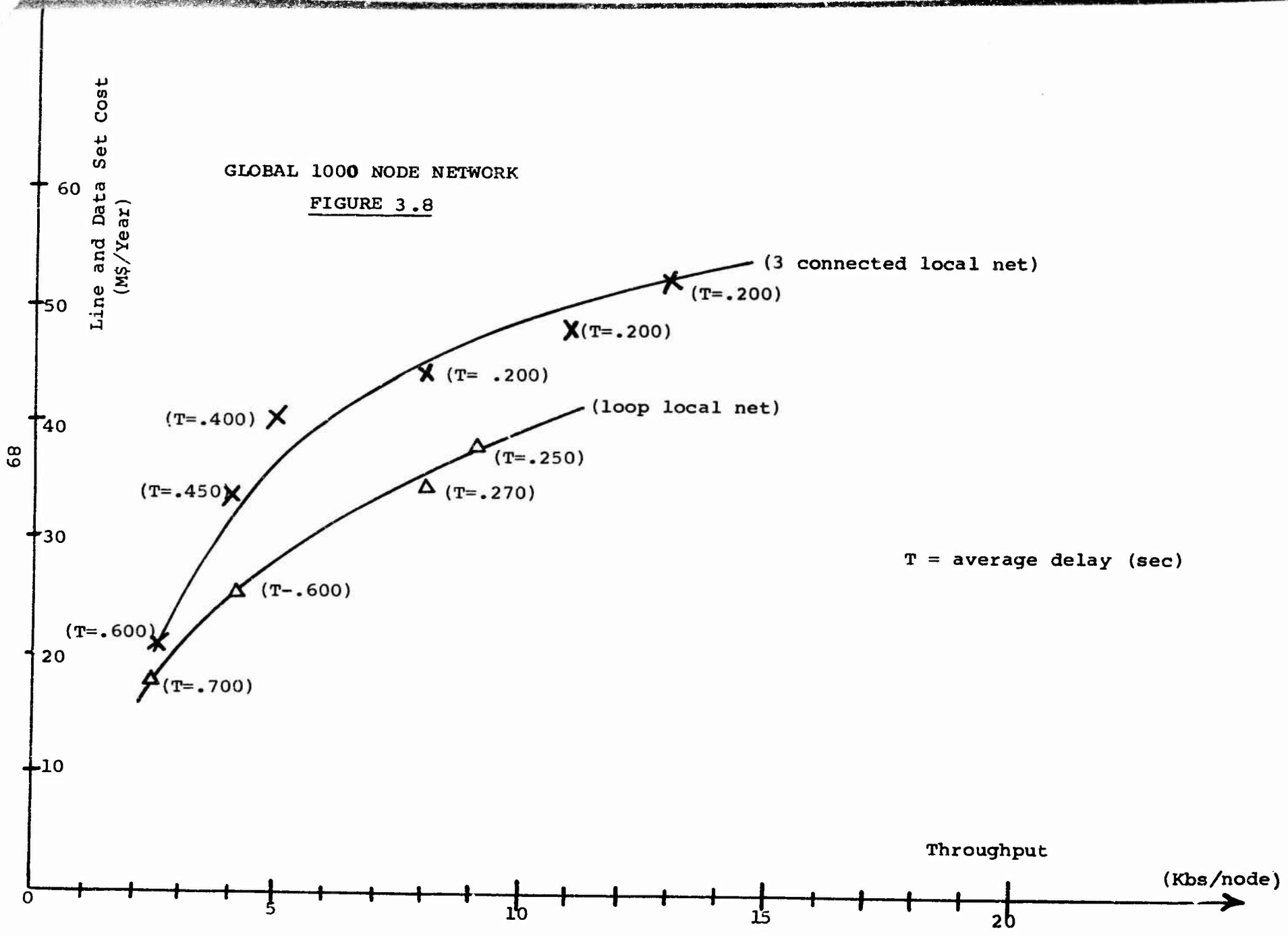
$$T_t = T_n + 2T_r + 2T_\ell$$

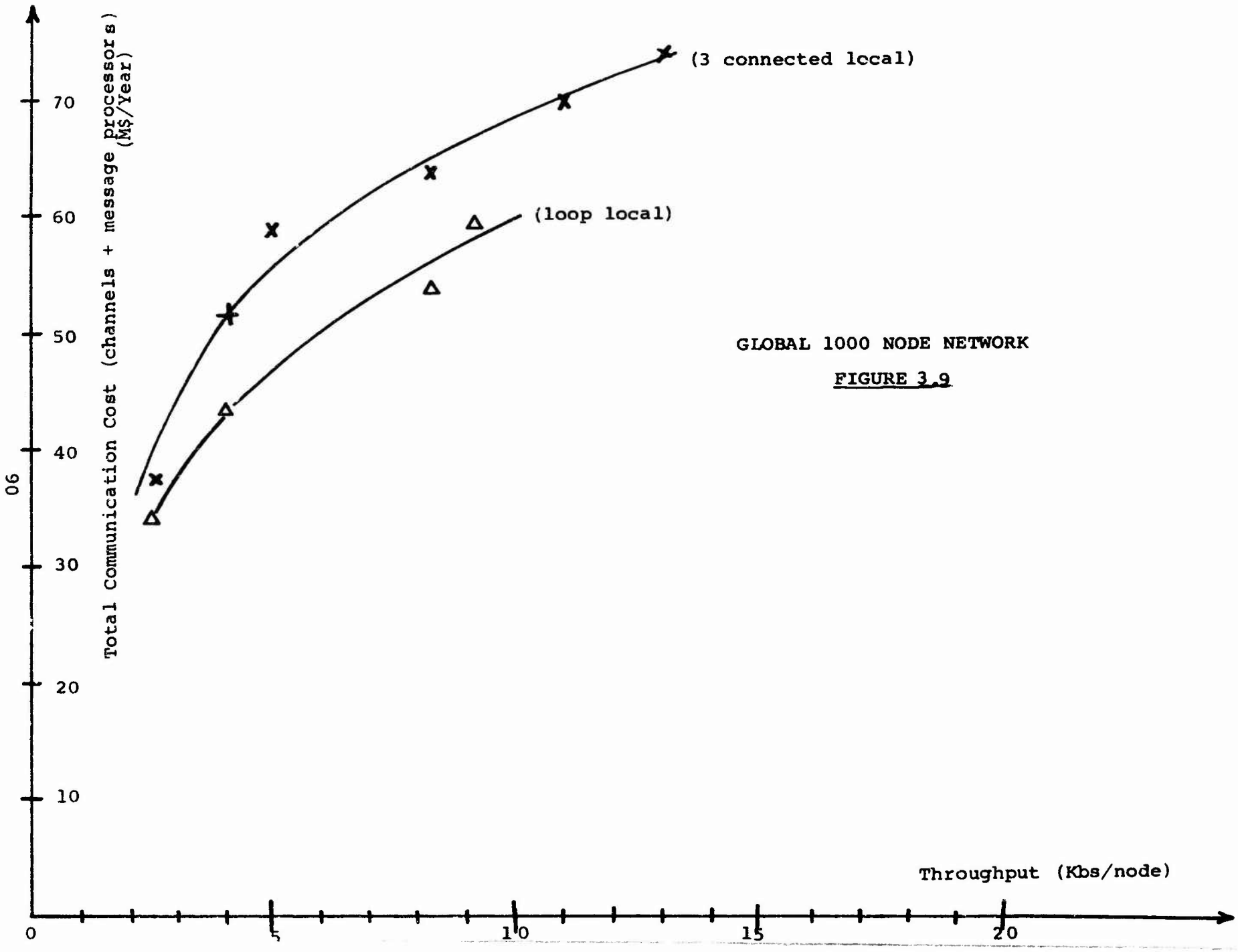
where T_n = national network delay

T_r = regional network delay

T_ℓ = local network delay.

Figure 3.8 shows channel cost and delay of the 1000 node net for both 3-connected and loop local net configurations. Figure 3.9 show the total communication cost, sum of channel and message processor costs.


GLOBAL 1000 NODE NETWORK
FIGURE 3.8
Line and Data Set Cost
(M$/Year)
60
50
40
30
20
10
68
(3 connected local net)
(T=.200)
(T=.200)
(T= .200)
(T=.400)
(T=.450)
(loop local net)
(T=.250)
(T=.270)
(T-.600)
(T=.600)
(T=.700)
T = average delay (sec)
Throughput
(Kbs/node)
0
5
10
15
20



GLOBAL 1000 NODE NETWORK

FIGURE 3.9

3.3 Reliability

To evaluate hierarchical network reliability, we make the assumption that two nodes in the same subnetwork can communicate with each other only through paths entirely contained in the subnetwork. Therefore, two node components of a subnetwork can be disconnected even if there is a connection path through the higher level network. This assumption is very realistic because, in a hierarchical routing implementation, the capability of sending local or regional traffic along paths external to the corresponding local or regional net, can be achieved only with considerable increase in complexity and overhead of the routing algorithm.

With the above assumption, the probability P_{nt} of the total network being disconnected is given by:

$$1 - P_{nt} = (1-P_{n\ell})^{100} \times (1-P_{nr})^{10} \times (1-P_{nn}) \times (1-P_{ex})^{110} \qquad (1)$$

where: $P_{n\ell}$ = probability of local net disconnected
P_{nr} = probability of regional net disconnected
P_{nn} = probability of national net disconnected
P_{ex} = probability of exchange node(or nodes) failure, which isolates the corresponding subnetwork.

Equation (1) can be rewritten as follows:

$$\log (1-P_{nt}) = 100 \log(1-P_{n\ell}) + 10 \log (1-P_{nr}) + \log(1-P_{nn}) + 110 \log (1-P_{ex}) \qquad (2)$$

If all disconnection probabilities (including P_{nt}) are small with respect to unity, Equation (1) becomes:

$$P_{nt} = 100\ P_{nl} + 10\ P_{nr} + P_{nn} + 110\ P_{ex} \qquad (3)$$

To evaluate F_{nt}, the fraction of disconnected node pairs, we make the simplifying (and conservative) assumption that whenever a subnetwork becomes disconnected, only one half of the nodes in the subnetwork can communicate, on the average, with the exchange node (or nodes). With such an assumption, if we let N be the number of nodes in the local net (in our case N=10) and a_l, a_r, a_n the number of noncommunicating node pairs resulting from the disconnection of a local, regional or national network respectively, we have:

$$a_l = \frac{N}{2}\left(N^3 - \frac{N}{2}\right) P_{nl} + N(N^3 - N)\ P_{ex}$$

$$a_r = \frac{N}{2}^2\left(N^3 - \frac{N^2}{2}\right) P_{nr} + N^2(N^3 - N^2)\ P_{ex}$$

$$a_n = \frac{N^3}{4} P_{nn} \qquad (4)$$

If we make the conservative assumption that the above contributions are statistically disjoint from each other, then we can sum them up and obtain the following expression for F_{nt}:

$$F_{nt} = 2P_n(1-P_n) + \frac{2}{N^6}\left(a_l N^2 + a_r N + a_n\right)$$

$$= 2P_n(1-P_n) + P_{nl} + P_{nr} + \frac{P_{nn}}{2} + 4P_{ex} \qquad (5)$$

where P_n is the node failure rate, and $2P_n(1-P_n)$ is the fraction of disconnected node pairs resulting from source and/or destination failures.

In order to evaluate P_{nt} and F_{nt} as from expressions (1) and (5), we need to know the network disconnection probability P_{nc} for the basic, 3-connected 10 node structure. The following results were obtained using NAC's reliability programs:

- $P_{link} = .02$; $P_{node} = .02 \rightarrow P_{nc} = 7 \cdot 10^{-4}$
- $P_{link} = .02$; $P_{node} << .02 \rightarrow P_{nc} = 8 \cdot 10^{-5}$

In the following, P_{nt} and F_{nt} are evaluated for a variety of network configurations which differ in:

(1) number of exchange nodes;

(2) redundancy in the exchange nodes;

(3) connectivity of the local network.

Figure 3.10 illustrates the various configurations.

(a) Only one exchange node; backups at all exchange nodes.

We have:

$P_n = 2 \times 10^{-2}$; $P_{nl} = 7 \times 10^{-4}$; $P_{nr} = P_{nn} = 8 \times 10^{-5}$; $P_{ex} = 4 \times 10^{-4}$

Thus:

$P_{nt} = 11.4 \times 10^{-2}$

$F_{nt} = 4.2 \times 10^{-2}$

(b) Three exchange nodes; backups at all nodes.

We have:

$P_n = 4 \times 10^{-4}$; $P_{nl} = 7 \times 10^{-4}$; $P_{nr} = P_{nn} = 8 \times 10^{-5}$; $P_{ex} = 6.4 \times 10^{-11}$

Thus:

$$P_{nt} = 8.8 \times 10^{-3}$$

$$F_{nt} = 1 \times 10^{-3}$$

(c) Three exchange nodes; backups at all exchange nodes.

We have:

$P_n = 1.4\text{x}10^{-2}$; $P_{n\lambda} = 7\text{x}10^{-4}$; $P_{nr} = P_{nn} = 8\text{x}10^{-5}$; $P_{ex} = 6.4\text{x}10^{-11}$

Thus:

$$P_{nt} = 7.8 \times 10^{-2}$$

$$F_{nt} = 2.8 \times 10^{-2}$$

(d) Three exchange nodes; no backups.

$P_n = 2\text{x}10^{-2}$; $P_{n\lambda} = 7\text{x}10^{-4}$; $P_{nr} = P_{nn} = 2.2\text{x}10^{-3}$; $P_{ex} = 8\text{x}10^{-6}$

Thus:

$$P_{nt} = 9.5 \times 10^{-2}$$

$$F_{nt} = 4.4 \times 10^{-2}$$

(e) Two exchange nodes; no backups.

We have:

$P_n = 2\text{x}10^{-2}$; $P_{n\lambda} = 7\text{x}10^{-4}$; $P_{nr} = P_{nn} = 1.3\text{x}10^{-3}$; $P_{ex} = 4\text{x}10^{-4}$

Thus:

$$P_{nt} = 12.8 \times 10^{-2}$$

$$F_{nt} = 4.4 \times 10^{-2}$$

(f) Only one exchange node; no backups;

We have:

$P_n = 2\text{x}10^{-2}$; $P_{n\lambda} = P_{nr} = P_{nn} = 7\text{x}10^{-4}$; $P_{ex} = 2 \times 10^{-2}$

Thus:

$$P_{nt} = 0.9$$

$$F_{nt} = 12 \times 10^{-2}$$

(g) Only one exchange node; backups at all exchanges; loop topology in local network.

We have:

$P_n = 2\text{x}10^{-2}$; $P_{n\ell} = 7.2\text{x}10^{-2}$; $P_{nr} = P_{nn} = 8\text{x}10^{-5}$; $P_{ex} = 4\text{x}10^{-4}$

Thus:

$$P_{nt} = 0.999$$

$$F_{nt} = 11.4 \times 10^{-2}$$

(h) Only one exchange node; backups at all exchanges; local network less than 3-connected.

We have:

$P_n = 2.10^{-2}$; $P_{n\ell} = 2.6\text{x}10^{-2}$; $P_{nr} = P_{nn} = 8.10^{-5}$; $P_{ex} = 4\text{x}10^{-4}$

Thus:

$$P_{nt} = 0.93$$

$$F_{nt} = 6.6\text{x}10^{-2}$$

(i) Only one exchange node; backups at all exchanges; local network less than 3-connected.

We have:

$P_n = 2\text{x}10^{-2}$; $P_{n\ell} = 1.8\text{x}10^{-2}$; $P_{nr} = P_{nn} = 8.10^{-5}$; $P_{ex} = 4\text{x}10^{-4}$

Thus:

$$P_{nt} = 0.84$$

$$F_{nt} = 5.8 \times 10^{-2}$$

(1) Only one exchange node; backups at all exchanges; local net less than 3-connected.

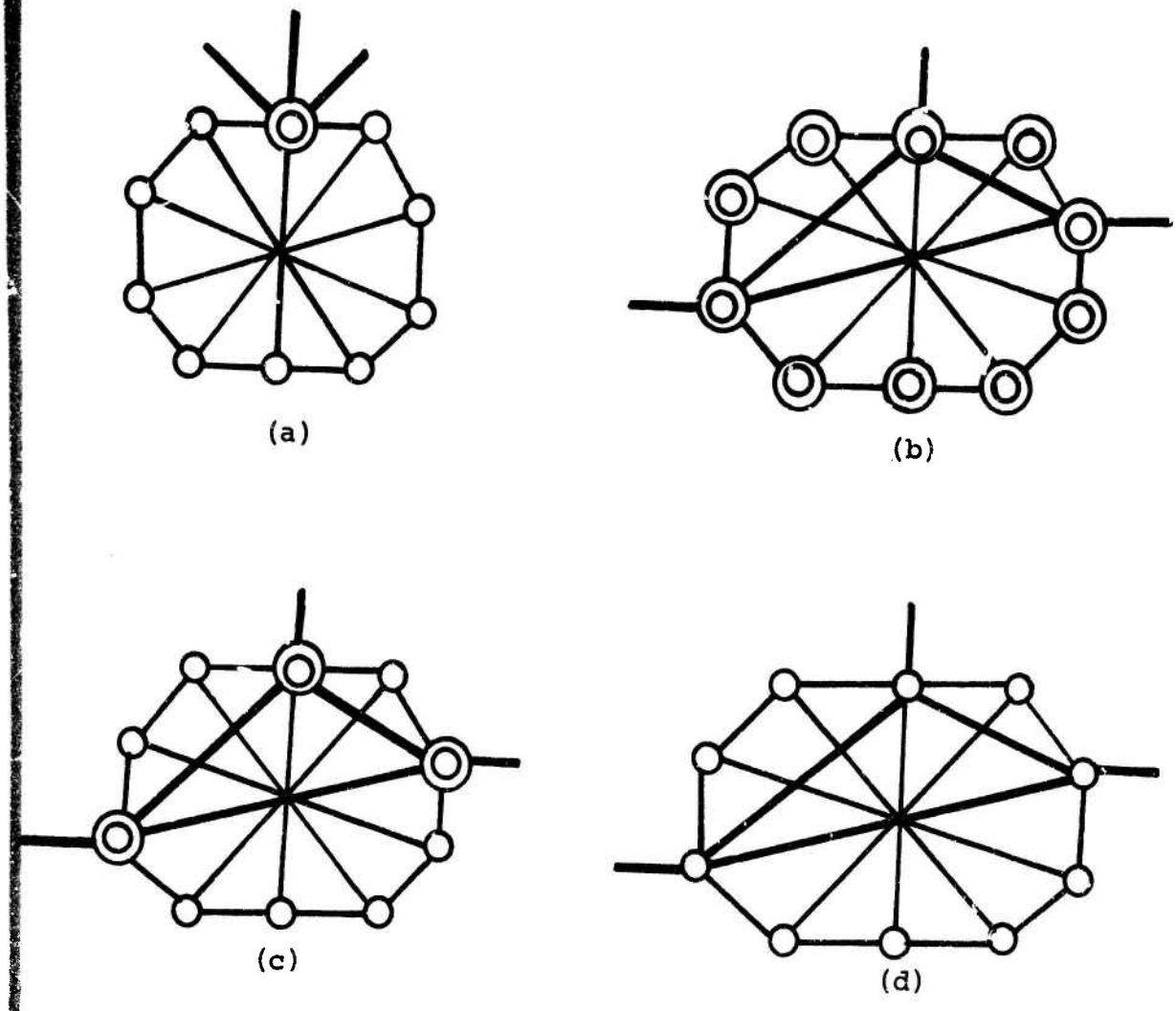
We have:

$P_n = 2 \times 10^{-2}$; $P_{n\ell} = 1 \times 10^{-2}$; $P_{nr} = P_{nn} = 8 \times 10^{-5}$; $P_{ex} = 4 \times 10^{-4}$

Thus:

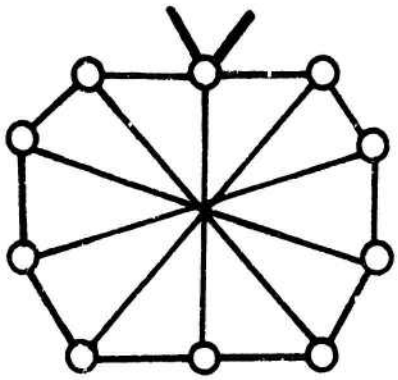
$$P_{nt} = 0.65$$

$$F_{nt} = 5 \times 10^{-2}$$

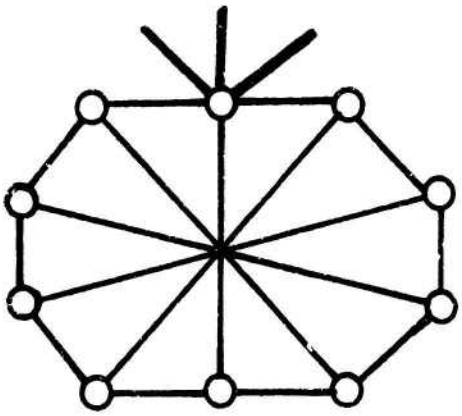


VARIOUS LOCAL NETWORK CONFIGURATIONS

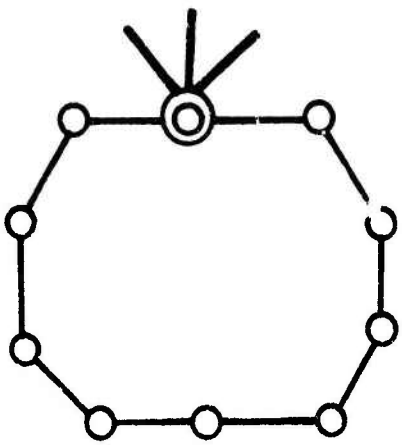
FIGURE 3.10



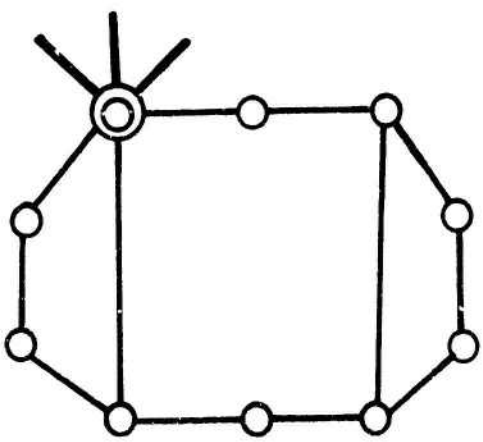
(e)



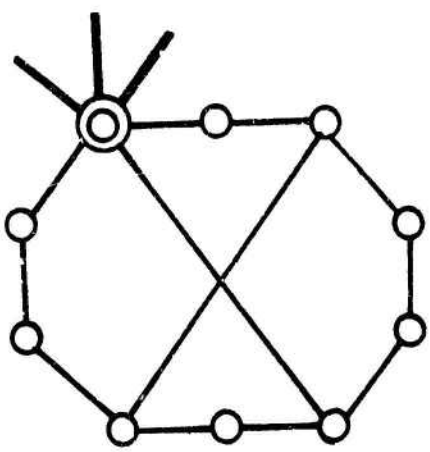
(f)



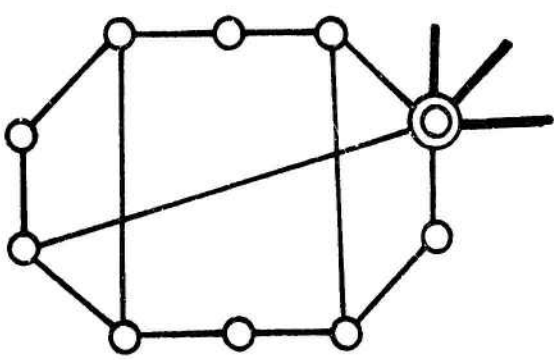
(g)



(h)



(i)



(j)

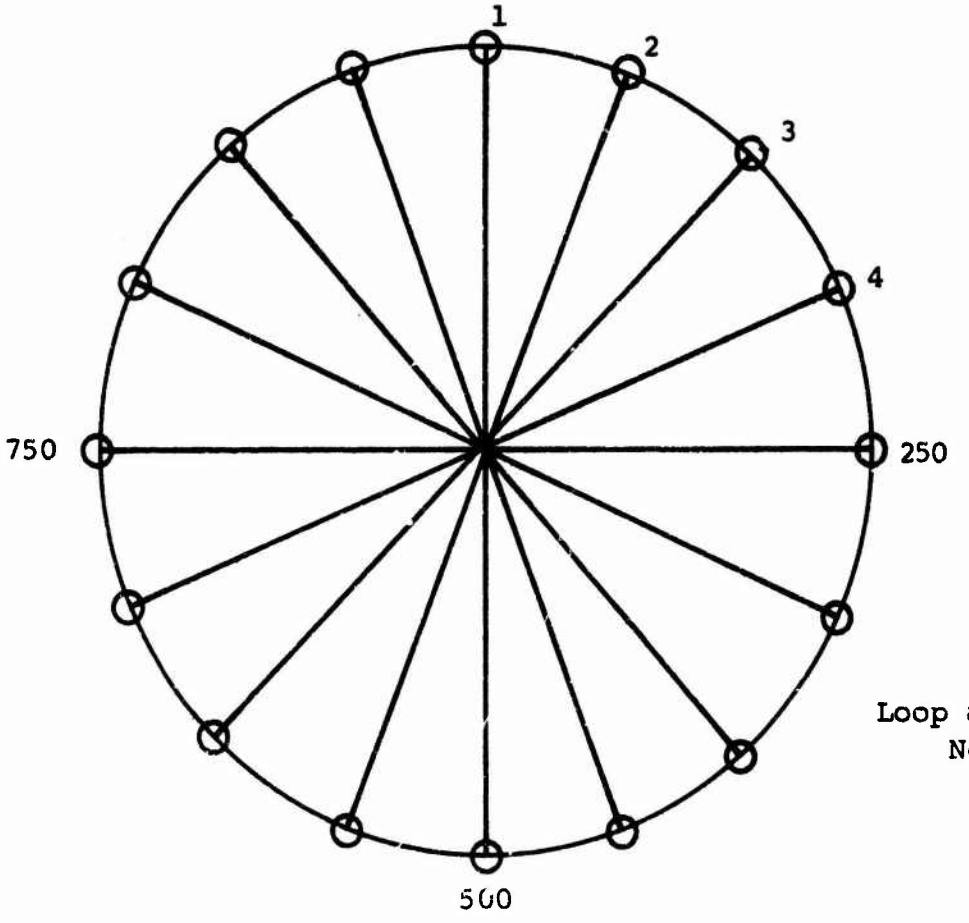
FIGURE 3.10 Concluded

4. NON-HIERARCHICAL NETWORK STRUCTURES

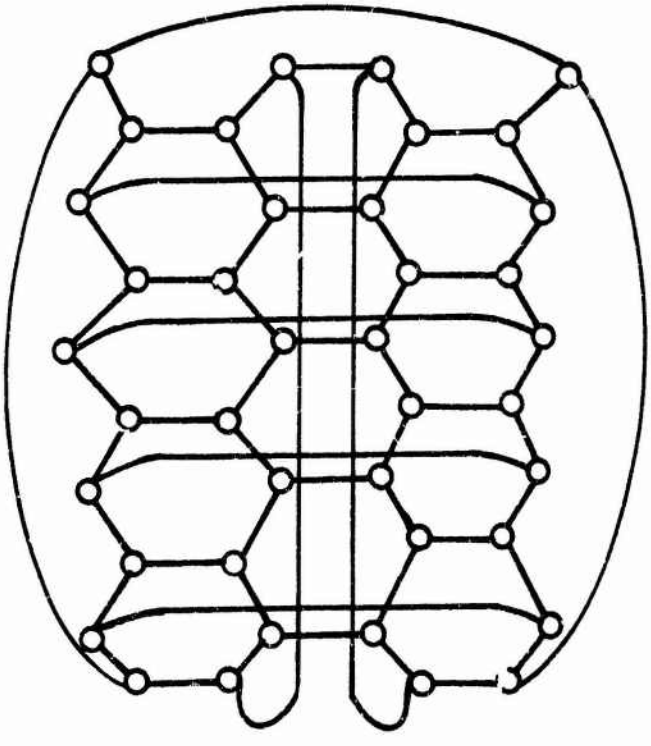
4.1 Introduction

In order to compare the hierarchical structure performance with that of non-hierarchical implementations, we analyzed two typical 3-connected non-hierarchical structures: (1) the "loop and star" network and (2) the "exagonal grid" network. The evaluation of such structures using the node locations of Figure 3.11 would be very cumbersome; therefore, a uniform distribution of the 1000 node locations over a 1000 x 2,500 miles rectangle was considered. This assumption generates error in the network cost evaluation, but provides exact answers for throughput, delay and reliability. Because of the homogeneous network structure, all nodes have the same importance, From the reliability point of view, therefore, since the cost of providing backups to all nodes is prohibitive, and no substantial improvement is gained by providing backups to only a subset of the nodes, we assume in the following that all nodes are non-redundant.

It is of interest to compare the behavior of average path length (i.e. the number of intermediate nodes on the "minimum link" path, averaged over all node pairs) as a function of network size, for hierarchical and non-hierarchical structures. If N is the number of nodes, then it is easy to see that the average path length is proportional to: N, for the loop and star network;



Loop and Star Network



Exagonal Grid Network

NON-HIERARCHICAL 3-CONNECTED STRUCTURES
FIGURE 3.11

$\sqrt{N}$ for the exagonal network; and, $\log_m N$, for a hierarchical structure with m nodes in each partition. The number of links is approximately the same for all structures; therefore, the average link traffic $\overline{f}$, given by

$$\overline{f} = \frac{\text{(input requirement)} \times \text{(average path length)}}{\text{number of links}}$$

is proportional to the average path length. Similarly, message delay is proportional to average path length. The above considerations show already an advantage of hierarchical structures, with respect to non hierarchical ones, for large network size.

4.2 The Loop and Star Network

Throughput, cost and delay analysis for the loop and star network can be easily performed by taking advantage of network symmetry. The following results were obtained:

- average path length = 125
- loop link traffic = 62.5 x R
- ray link traffic = .5 x R

where R is the throughput per node.

The above results assume "minimum link" routing between node pairs.

A channel capacity allocation with 230.4 Kbs on the loop links and 9.6 Kbs on the ray links was first considered. The

results follow:

Channel Cost = 55 M$/year

Throughput = 3 Kbs/node

Delay = .300 sec.

Next, 1,544 Kbs channels were assigned to loop links, and 19.2 Kbs to ray lirks. The results follow:

Channel Cost = 120 M$/year

Throughput = 20 Kbs/node

Delay = .050 sec.

Network reliability was evaluated by counting only the disconnections produced by 3 and 4 element failures, thus obtaining an optimistic estimate. The results follow:

$P_{nt} = .45$

$F_{nt} = 17 \times 10^{-2}$

where P_{nt} is the probability of network disconnected and F_{nt} is the average fraction of disconnected node pairs.

4.3 The Exagonal Grid Network

In the exagonal network case, the symmetrical structure allows a straightforward evaluation of cost, throughput and delay. The following results were obtained, assuming minimum link routing:

- average path length = 20
- link traffic = 6.6 x R

where R = throughput per node.

With uniform 50 Kbs capacity assugnment, the following results were obtained:

Channel Cost = 36.5 M$/year

Throughput = 6 Kbs/node

Delay = .200 sec.

With uniform 230.4 Kbs capacity assignment, the following results were obtained

Channel Cost = 87 M$/year

Throughput = 28 Kbs/node

Delay = .045 sec.

An optimistic upper bound on network reliability was evaluated by considering all the possible ways of obtaining disconnected node components with one, two or three nodes (the probability of higher component disconnections was assumed to be negligible. The results follow:

$$P_{nt} = 7 \times 10^{-2}$$

$$F_{nt} = 4 \times 10^{-2}$$

where P_{nt} = probability of network disconnected; F_{nt} = fraction of disconnected node pairs.

5. DISCUSSION

The results of the previous sections provide a very important insight into the design and performance of large, distributed communication networks. Using such results, we now discuss the very vital issues of topological structure, reliability and cost-throughput trends for large networks.

5.1 Hierarchical and Non-Hierarchical Structures

Considering the two non-hierarchical structures proposed in Section 4, the loop and star network appears to be much more expensive, and much less reliable, than the exagonal network, and therefore is eliminated from further consideration. Next, if we compare exagonal and hierarchical network results, we notice that the costs for a given throughput are about the same while reliability performance is slightly better in the exagonal network. The following considerations, however, make the hierarchical structure appear more attractive for large networks:

- the hierarchical structure can utilize, in each subnetwork a routing algorithm similar to the present ARPANET algorithm. In the exagonal structure, a completely new algorithm should be developed.

- the hierarchical structure offers great flexibility in the design. In fact, a large gamut of cost-throughput solutions

can be obtained by gradually increasing some of the channel capacities at various hierarchical levels (see Figure 3.8). In the exagonal structure, on the other hand, all nodes and links have the same "importance"; therefore, a selective capacity increase is meaningless, and large gaps between solutions are likely to ccur. For example, in the 1000 node case discussed in Section 4.3, no intermediate solution exists between the 6 Kbs/node and the 28 Kbs/node solution.

- The hierarchical structure offers more control on network reliability. In fact, by introducing redundancy in some of the nodes and by varying the connectivity of some of the hierarchical levels (see Section 3.3), cost-reliability tradeoff can be adjusted to the specific application under study. Furthermore, if the network contains nodes with important resources, or nodes with special control tasks, the access to such nodes can be improved by including them in higher hierarchical levels.

- The hierarchical structure can easily adapt to network growth. In fact, if the introduction of new nodes, or external traffic increase in some sections of the network configuration, only a few subnetworks of the hierarchical structure have to be modified. In non-hierarchical structures, on the other hand,

the network reconfiguration would probably be more complex and not as efficient.

- The hierarchical structure can achieve better system economics using different communication schemes in different hierarchical levels. For instance, local networks could be implemented with multidrop lines, multiplexed lines, packet radio communications, etc. The national network, on the other hand, could use satellite links.

- The average path length is proportional to logN in the hierarchical case, and to $\sqrt{N}$ in the non-hierarchical one (where N is the number of nodes). It is conceivable, therefore, that beyond some value of N the hierarchical structure is not only more efficient, but also less costly than the non-hierarchical one.

5.2 Reliability of Large Networks

From the results of Section 3.3 it can be observed that F_{nt} is in most cases very close to the lower bound $2P_n(1-P_n)$, while P_{nt} tends to become very large as soon as exchange node redundancy or 3-connectivity are relaxed. This behavior is typical of very large networks; for example, if we analyze a 100,000 node network with 5 ten node hierarchical levels (each level 3-connected; 3 exchange nodes; all nodes perfectly reliable) we obtain:

$$P_{nt} = .6$$

$$F_{nt} = 4 \times 10^{-4}$$

Clearly, more than 3-connectivity is required to obtain an acceptable value of P_{nt}, while F_{nt} is extremely low with a 3-connected configuration.

In general, for a hierarchical structure with m nodes in each subnetwork, we have from Equations (3) and (5), the following relations between P_{nt}, F_{nt} and N:

$$\begin{aligned} P_{nt} &\cong K_1 \frac{N}{m} \times P_s \\ F_{nt} &\cong K_2 (\log_m N) \times P_s + P_n(1-P_n) \end{aligned} \tag{6}$$

where K_1 and K_2 are proportionality coefficients independent from N, P_s is the disconnection probability of the basic m node subnetwork and P_n is the average node failure rate. It is obvious therefore that P_{nt} is much more difficult to control than F_{nt} for large N.

In the design of large networks, therefore, the selection of either P_{nt} or F_{nt} as the reliability criterion leads to different solutions for the basic topological structure. In the 1000 node case for example, if $P_{nt} \leq 0.15$ is required, then 3-connectivity is necessary; if on the other hand $F_{nt} \leq .06$ is

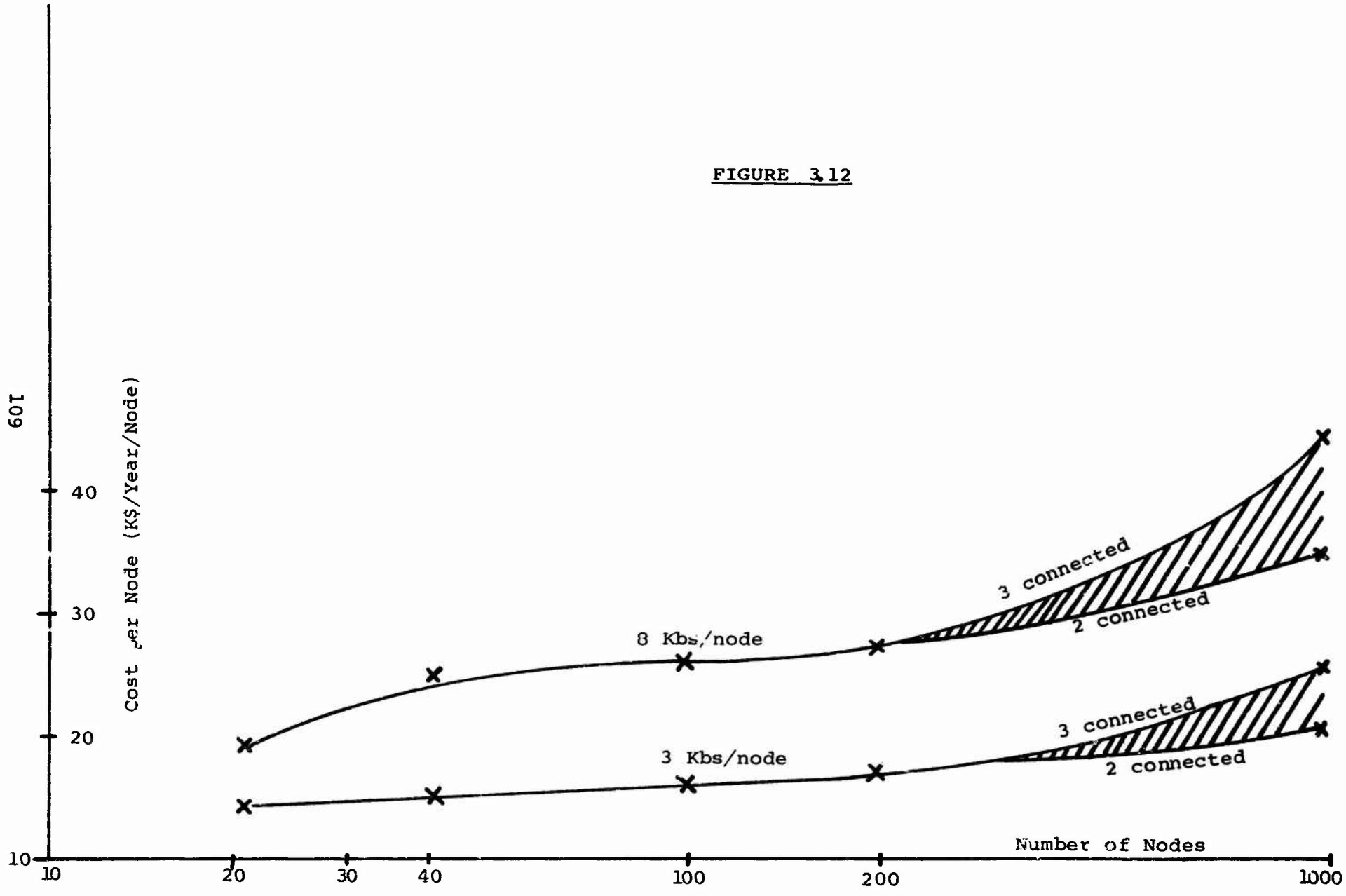
required, local network topologies less then 3-connected lead to acceptable, more economical designs. More specifically, in the latter case a good, low-cost 1000 node design can be obtained with: 2 exchange nodes; no backups; local network less than 3 connected.

The fact that F_{nt} is close to the lower bound $2P_n(1-P_n)$ for most of the cases discussed in Section 3.3 shows that substantial improvement of F_{nt} is obtained only by improving the reliability of the message processors. Work is presently under way in this direction.

5.3 Cost and Throughput Trends

The diagram in Figure 3.12 displays line and modem cost per node versus network size, for two different values of throughput The data for N=20, 40, 100 and 200 were obtained from previous NAC studies. The shadowed area represents the cost of networks with local connectivity ranging from 2 to 3. The cost for N=1000 seems to be slightly higher than the trend displayed for N up to 200. It should be remembered however that: (1) the cost estimate for N=1000 is not very precise; (2) the cost for N $\leq$ 200 was minimized using branch exchange procedures, while the cost for N=1000 is just a feasible cost. Thus, we can expect that the optimized network cost for N=1000 would be lower than the feasible cost, and could follow very closely the trend already established for N up to 200.

FIGURE 3.12


109
Cost per Node (K$/Year/Node)
40
30
20
10
8 Kbs/node
3 Kbs/node
3 connected
2 connected
3 connected
2 connected
Number of Nodes
10
20
30
40
100
200
1000

6. IMPLICATIONS FOR FUTURE RESEARCH

The results contained in this report, together with those of previous reports, establish the feasibility, in terms of design techniques, cost, delay and reliability, of very large packet switched network design. Future steps in the research will be: (1) optimization of network design; (2) performance evaluation; (3) routing and flow control; (4) use of different communication techniques at different hierarchical levels. Some of the open areas are elaborated in the following.

- Optimal Design

  The design of a hierarchical network requires: selection of number of hierarchical levels and of number of "nodes" for each level; determination of node partitions (on the basis of geographical distance, node requirements, etc.); separate minimum cost design for each partition and hierarchical level; combination of the partial designs into the global design. Low cost designs can be obtained with iterative procedure, in which an initial configuration is successively improved, by properly modifying node partitions, local topologies, interconnections between different hierarchical levels, etc., until no more improvement is possible. One of the bottlenecks of the procedures is the

local minimum cost network design, which must satisfy both traffic and reliability constraints. The present branch exchange techniques are inadequate; faster, and more efficient methods must be developed. In addition, the use of the interactive graphics programs presently being developed at NAC will be very valuable in the application of the above iterative procedure by simplifying the input of network data, the determination of successive node partitions and the modification of various local topologies.

- Performance Evaluation

The exact evaluation of throughput, delay and reliability for a 1000 node network requires a prohibitive computation time and memory space if performed with the present methods. This is not so critical for the network design in which the approximate expressions of throughput, delay and reliability derived in Section 3 are probably sufficient. For the final configuration, however, a more precise performance evaluation is required; therefore, new and efficient methods for large network analysis must be developed.

- Routing and Flow Control

The traffic within each subnetwork can be routed and controlled with the present ARPANET techniques. However, proper

modification must be introduced in order to direct the traffic to external destinations. In addition, a multilevel flow control could be implemented, in order to obtain a more efficient control of the traffic load in each hierarchical level.

- Hybrid Communication Implementations

The hierarchical structure allows within certain limits the use of different system implementations at different hierarchical levels. This feature can be exploited in order to obtain a more economical and efficient system. Possible configurations might include: packet radio techniques at the local level; packet switching techniques at the regional level; satellite broadcast techniques at the national level. It is of interest to investigate feasibility and economics of such hybrid implementations.

4. TERMINAL ORIENTED NETWORK COST AND PERFORMANCE--PART I

1. INTRODUCTION

The ARPANET was originally conceived as primarily a computer communication system. As the network was implemented, it quickly became obvious that terminal to computer traffic was playing a significant role in network usage. The TIP (Terminal Interface Processor) was then developed to provide ARPANET access for terminals without dependence on a Host computer.

As the Network has grown, the TIP has assumed a vital role in network communications. The TIP was originally conceived of as a device providing mainly dialup services. However, several leased lines are now connected to the TIP and experiments are beginning at NAC to multiplex several low speed terminals over a single voice grade line.

It is becoming evident that effective widespread usage of the ARAPNET approach within the Defense Department could involve a system with possibly hundreds of Hosts and tens of thousands of low speed terminals. Economical use of a network will thus depend on cost effective terminal access as well as efficient computer to computer communications. The Packet Radio project provides one extremely promising approach to the terminal access problem. However, in order to evaluate the merits of

Packet Radio and to provide alternatives in areas where packet radio systems are not desirable, it is necessary to investigate other suitable access schemes.

In this chapter, we discuss the first results of an ongoing study to investigate cost-performance tradeoffs as a function of the number of terminals within the system. The major effort during the present reporting period has been the construction of appropriate design tools with which to study the problem.

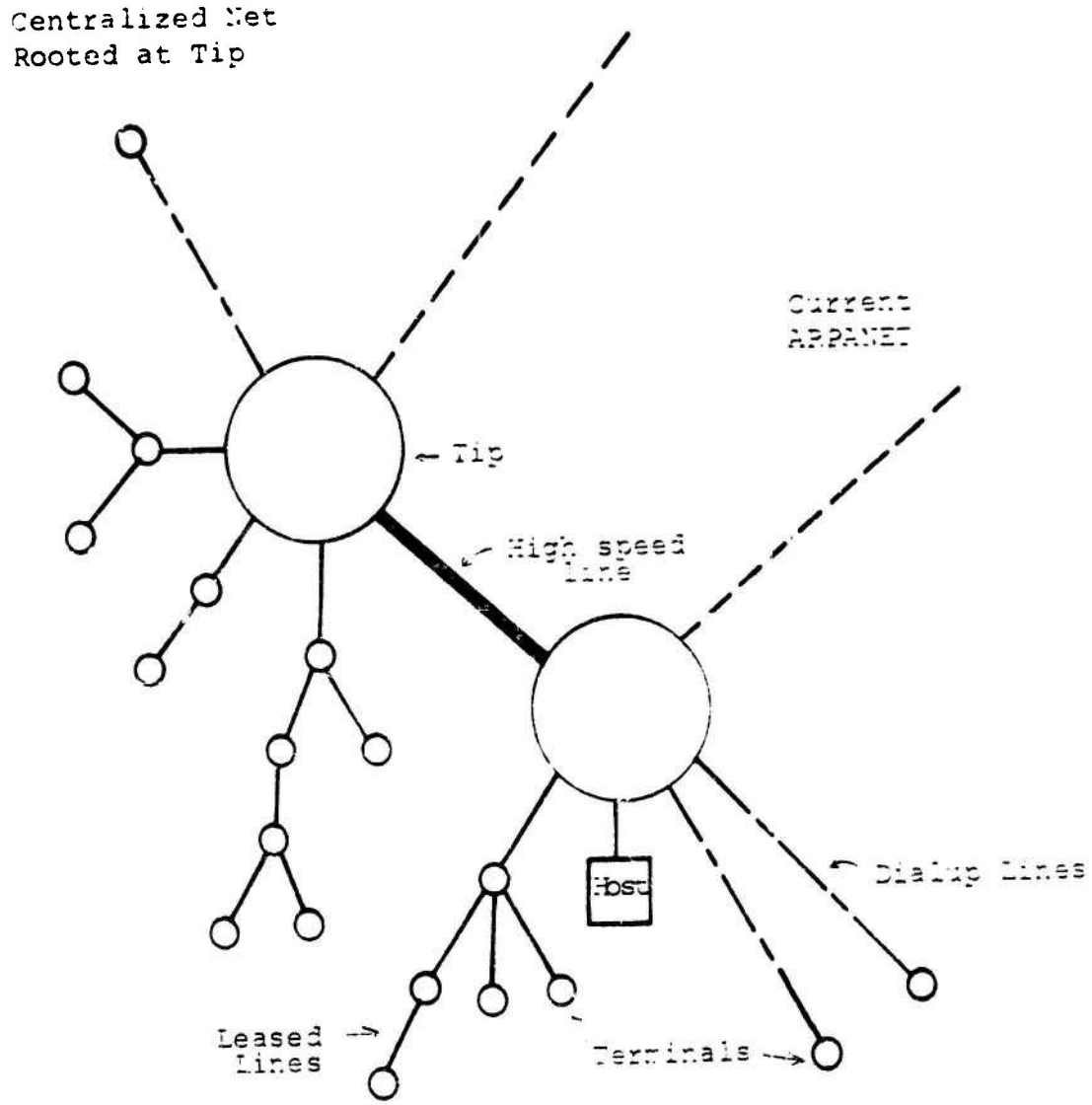
We consider the problem of extending the ARPANET by using TIPs as the roots of centralized networks, composed of multi-dropped leased voice grade lines or dialup lines. Thus, the size and usefulness of the network can be increased over an order of magnitude without a corresponding increase in cost. It is also thus possible to open the net to the large class of small users who do not have a level of traffic large enough to warrant a TIP or high speed line.

2. NETWORK ARCHITECTURE AND PROTOCOL

A simplified diagram of the network architecture is shown in Figure 4.1. In it, each terminal communicates with others and with the ARPANET hosts through the TIP at the root of the centralized network of which it is a part. Routing of traffic among the TIPs is handled by the same means currently in use in the ARPANET. Communication between a terminal and its associated TIP can be handled in one of two ways.

The TIPs can poll each line in their associated networks and information is then transferred using a protocol for polled multidropped networks. For example, lines may be polled sequentially until a positive acknowledgment is received at which point polling of that line stops until the incoming message is received by the TIP and transmitted to its destination. Similarly, when an outbound message from TIP to terminal is ready for transmission, polling of that line is interrupted for transmission of the outbound message.

While polled centralized data communications systems are currently popular and widely used, they do possess a disadvantage which could be particularly large in this system. The disadvantage is that polling carries with it an associated overhead which is directly proportional to the number of terminals on each line



NETWORK ARCHITECTURE

FIGURE 4.1

and the ratio of the length of the poll and the length of the messages sent. By its very nature, this system is composed of a large number of terminals, each with relatively low traffic levels. Thus, the polling overhead as a percentage of total capacity will be high.

This leads us to the possibility of using a network protocol similar to that of the Packet Radio networks where each terminal transmits to the root whenever it has a message and two messages arriving at the root at the same time interfere with one another and both must retransmit. Thus, all the terminals on the same multidrop line share a common "channel". It has been shown that such a protocol will yield an effective line utilization of $\frac{1}{2e}$ which in a system such as this one, where there is a large overhead associated with polling, may already represent an improvement. It is possible, however, with a simple modification to the protocol, to increase the line utilization significantly, indeed, to the point where it is near 1. Instead of broadcasting a packet of information when it has a message, each terminal transmits its identifier on a pilot frequency. Upon reception of this identifier, the TIP sends a selection sequence (most likely the terminal I.D.) which authorizes the terminal to transmit. Thus, the overhead of retransmitting packets which interfere with

one another is reduced to one of retransmitting the identifiers when they are simultaneoulsy received. Since the identifiers are in general much shorter than packets, there is less to retransmit. In addition, the identifiers are far less likely to overlap in time. Essentially, what we accomplish by implementing such a scheme instead of polling is to eliminate the overhead associated with polling and getting a negative response. This is clearly an improvement, and in this case, a large one.

3. SCOPE OF PROJECTED OVERALL ANALYSIS

The above problem has many facets which interrelate strongly with one another. The first is the establishment of an appropriate data base. We will consider a prototype composed of users located at population centers in a major metropolitan area with traffic levels proportional to population size. The constant of proportionality will vary in order to consider various tradeoffs among network components.

Related to this is the question of how many TIPs should be used in the design. There is a minimum number of TIPs which must be used given any level of traffic as each TIP has a capacity asscciated with it. There is also a capacity associated with the high speed lines connecting the TIPs. Thus, the total traffic in a single centralized network is limited and the system will in general contain at least C_{min} centers:

$$C_{min} = \left[\frac{T}{T_{max}} \right]$$

where T is the total traffic in the system and T_{max} is the maximum allowable traffic level in any one centralized network.

The actual numbers of TIPs, however, will in general be larger than this since additional centers will reduce line charges by concentrating many low speed lines into a single high speed line.

As the level of traffic rises, however, coverage by the TIPs will increase and savings obtained by placing additional TIPs in the area will decrease. Thus, the determination of how the number of TIPs grows with increasing traffic must be made in the context of the overall problem and will be investigated in later studies.

Another closely related question is where the TIPs should be located. A solution to this facet of the problem must consider the tradeoff between placing TIPs to minimize the cost of the voice grade lines comprising the centralized networks and placing them to minimize the cost of the high speed lines comprosing the interconnecting network among TIPs. In general, these goals conflict and a balance must be struck between them to minimize overall system cost.

Another question is which TIP should each terminal be associated with. In forming a partition of the terminals with respect to the TIPs, we must consider not only the minimization of costs in the centralized multidrop networks but also the capacity restrictions on the TIP and their associated high speed lines.

Closely related to this problem is the layout of the multidrop lines themselves, that is, how terminals should be interconnected along the path to their associated TIP. As we will see, this

problem is basically one of finding constrained minimal spanning trees. The constraint in forming the spanning trees is that the total traffic at nodes connected to a TIP along the same line not exceed some specified limit which will guarantee acceptable network performance. This maximum is a function of line speed, number of terminals on each line, and the network protocol. The calculation of this function is itself a significant problem.

As we have already mentioned, all the above questions are interrelated facets of the same problem and can meaningfully be answered only within the context of the problem as a whole. We will show, however, that they can be solved to a great extent sequentially if proper use is taken at each stage of the solution.

We will begin by considering a method for the solution of the problem of laying out the multidrop lines, which is the topic of the remainder of this discussion. In solving this problem we assume that we already know the number and locations of the TIPs, the actual values of the traffic constraints on the lines, and which TIP each terminal is associated with. The method presented is very flexible and will, in fact, produce a good solution to the layout problem over a wide range of solutions to the other problems. It is also very fast and can therefore be extended to solve the layout problem as a part of the solution of some of

the others. It will also be shown in a later section that this method can be further extended to additionally yield a solution to the problem of associating terminals with TIPs without appreciably increasing its running time.

4. LINE LAYOUT OPTIMIZATION

The problem is to produce a low cost tree connecting the terminals to the central node subject to constraints on the number of terminals and total traffic in each multidrop line. Such constraints are necessary to ensure that line capacities are not exceeded and the delay time for a response along any given line is kept within reasonable bounds.

In the discussion that follows we will speak of generating trees of minimum length. We are actually concerned with generating trees of minimum cost. Line charges are not, in general directly proportional to length, particularly when interstate lines are charged at a lower rate than intrastate lines. We will, however, consider length and cost to be proportional as it is usually clear from the context of the problem whether a given multidrop line will be interstate or intrastate and the "lengths" we assign to branches can in fact be their costs.

Many algorithms have been proposed for the solution of this problem. These algorithms can be divided into two disjoint classes which are fundamentally different in their approach and results. The first class contains all the algorithms which approach the problem from the point of view of integer programming. There are several drawbacks to such an approach. First, it is

very difficult to implement the non-linear constraints which often arise in practical communications networks using such an approach. One is usually forced to approximate such constraints by tighter linear constraints with resulting degradation of the solution. Second, it is very hard to alter the constraints during the solution. It is often desirable to do so in order to perturb the solution trading cost versus performance. Finally, such methods possess no polynomial computational bound and one usually must terminate execution before an optimal solution is found for any reasonable problem.

The second class, which is the one we will deal with here, contains heuristic algorithms which overcome the first two objections mentioned above and which for reasonable sized problems will usually produce better solutions than the algorithms in the first class do in a comparable amount of time. The basic concept underlying most of these algorithms is the same. In each case, the nodes are initially placed in separate components and pairs of components are then joined by the shortest arc in the cut separating them. Rosenstiehl (1967) proved that algorithms of this form will always generate spanning trees of minimum total length (MSTs) if they are not restricted in their choice of cuts. Unfortunately, each of these algorithms must consider the

constraints during the course of execution and therefore their choice of cuts is restricted. Thus, unless no pair of components which these algorithms would otherwise merge violates a constraint, they do not in general generate an MST or even the minimal cost tree satisfying the constraints. The question as to whether an algorithm with a polynomial bound exists to solve this problem optimally, is as yet undecided. Experience with these algorithms has shown, however, that they do generate good solutions, within a few percent of the optimum in most cases. Furthermore, since the nodes on each multidrop line satisfy the constraints, the multidrop line is in fact an MST on these nodes and the central node.

These algorithms are powerful in their ability to treat problems with a large variety of constraints which may differ from one another in functional form. We require only that it be possible at any stage of the algorithm to connect each component directly to the central node and obtain a feasible solution. In the following discussion, we associate with each node (or component) i a finite vector, $\overline{V}_i$, which contains information about the node (or component) relevant to the calculation of the constraints. The value of each constraint at node i is then considered to be some $F(\overline{V}_i)$, where the only constraints on F are the above one, that it be computable from the information

contained in $\overline{V}_i$, and that F be computable on the component formed by joining nodes i and j given $\overline{V}_i$ and $\overline{V}_j$.

Note that it is not necessary for the graph from which the spanning tree is to be chosen to be complete, only that it contain an arc linking each node directly to the central node. This concept is particularly important when the number of nodes is so large that the number of arcs in a complete graph on these nodes exceeds the memory capacity of the computer being used. It is possible, without significantly degrading the solution obtained, to treat such problems within the context of a relatively sparse graph where each node is connected to a small number of its nearest neighbors and to the central node. Important savings in core requirements and running time can thus be obtained.

We now present a brief description of several of the most widely accepted heuristic algorithms. In each case, before joining two components, we check to see if having the nodes in both components on the same multidrop line violates any constraints. If so, the algorithm does not join the components and proceeds to consider the next candidate pair.

Prim's Algorithm

Initially, one node is in the spanning tree. At each stage, the node whose distance to any node already in the tree is minimal is brought into the tree.

Kruskal's Algorithm

Initially, each node is in a separate component. At each stage, the shortest arc connecting nodes in different components is found and these components are joined by that arc.

Esau-Williams Algorithm

Define a tradeoff function, t_{ij}, as the length of the arc connecting node i to the central node minus the length of the arc connecting node i to node j if an arc exists between nodes i and j. At each stage we find the largest t_{ij} and bring the arc (i,j) into the spanning tree.

VAM Algorithm

Define d_i as the distance between node i and its nearest feasible neighbor (i.e., its nearest neighbor which can be placed on the same multidrop line as node i). Define b_i as the distance between node i and its second nearest feasible neighbor. Define a tradeoff function, t_i, as $b_i - d_i$. At each stage, find the largest t_i and join node i to its nearest feasible neighbor and treat the resulting component as a node.

Each of these algorithms starts with the nodes in separate components and subsequently joins pairs of components. They differ only in the order in which they consider joining components. Since the grouping of a given set of nodes into one component

(i.e., placing them on the same multidrop line) restricts subsequent merging with other components because of the constraints, these algorithms, in general, yield different solutions. As an example, consider the application of each of the above algorithms to the graph shown in Figure 4.4. Prim's algorithm would

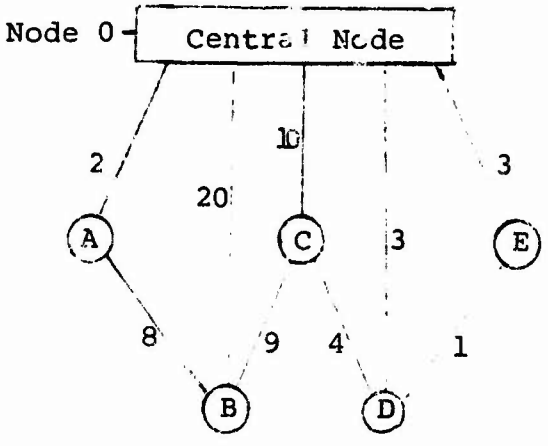


FIGURE 4.2

consider arc (A,O) first. Kruskal's algorithm would consider arc (D,E) first. Esau-Williams' algorithm would consider arc (B,A) first. The VAM algorithm would consider arc (C,B) first. Thus each algorithm starts by forming a different component and could in general yield a different solution from any of the others. In each case, however, the arc chosen links some node to its nearest neighbor, subject to the constraints. The reason the algorithms choose different arcs is that they consider the

nodes in a different order. They implicitly associate weights with the nodes and use these weights to decide the order components are to be considered for merging. Using this concept, it is possible to develop one algorithm which will implement all the above algorithms and many others of the same type. The following procedure will implement this class of algorithms given a set of rules, which we call w-rules, for initializing and updating the node weights.

Unified Algorithm

Definitions of variables

w_i - weight associated with component i

C_i - component containing node i

S_i - number of nodes in component i

$\overline{V}_i$ - vector containing information relevant to the calculation of constraints on component i

t_{ij} - tradeoff function associated with arc (i,j)

d_{ij} - length of arc (i,j)

Step 0: Initialize the w_i $i = 1, 2, \ldots, n$ using the appropriate w-rule. Initialize the $\overline{V}_{ij}$. Set $t_{ij} \leftarrow d_{ij} - w_i$ when d_{ij} exists and $C_i \cup C_j$ does not violate any constraints.

Set $C_i \leftarrow i$ $\quad i = 1, 2, \ldots, n$

Set $S_i \leftarrow 1$ $\quad i = 1, 2, \ldots, n$

Go to Step 1.

Step 1: Find $t_{i^*j^*} = \min_{\substack{i,j \\ C_i \neq C_j}} t_{ij}$

If $t_{i^*j^*} = \infty$, terminate, otherwise go to Step 2.

Step 2: Evaluate the constraints on $C_{i^*} \cup C_{j^*}$. If any are violated, set $t_{i^*j^*} \to \infty$ and go to Step 1. Otherwise go to Step 3.

Step 3: Add arc (i^*, j^*) to the spanning tree. If $S_{i^*} \leq S_{j^*}$, set $K \leftarrow i^*$, $\ell \leftarrow j^*$ otherwise set $K \leftarrow j^*$, $\ell \leftarrow i^*$. Set $S_K \leftarrow S_{i^*} + S_{j^*}$. Set $C_i \leftarrow K \ \forall_i \in C_\ell$. Re-evaluate the $\overline{V}_{\ell j}$. Update w_i using the appropriate w-rule and re-evaluate the t_{ij}. Go to Step 1.

The specific w-rules for implementing the algorithms mentioned earlier are given in Table 4.1.

TABLE 4.1

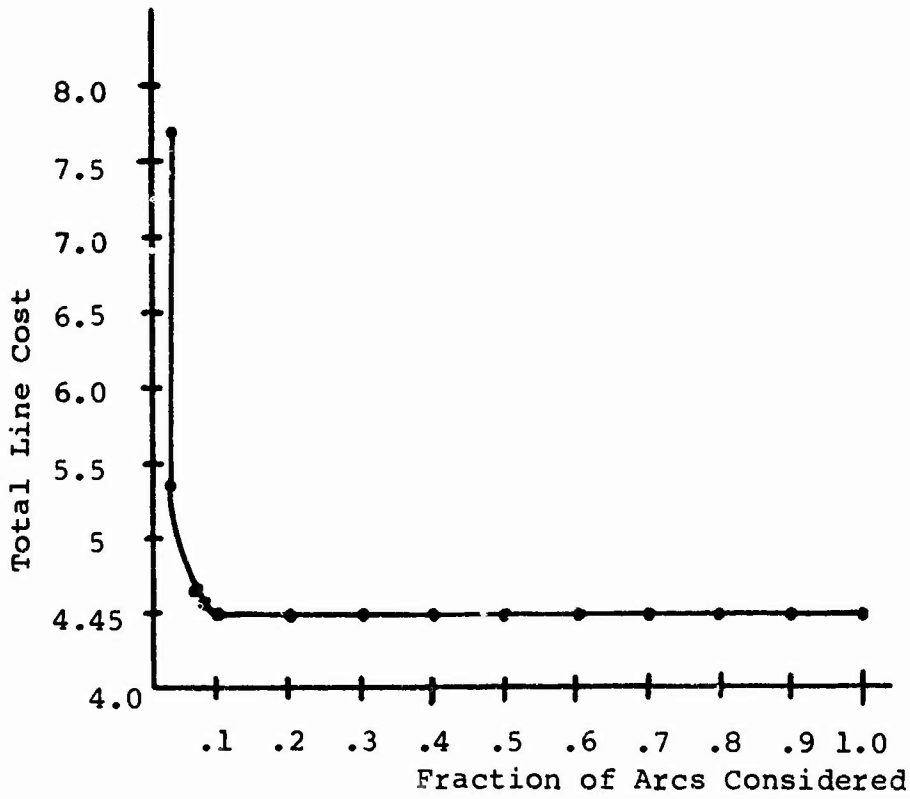
Algorithm	Initialization	Update when Arc(i,j) is brought in
Prim	$w_1 = 0$ $w_i = -\infty \quad i=2,\ldots n$	$w_j \leftarrow 0$
Kruskal	$w_i = 0 \quad i=1,\ldots n$	none
Esau-Williams	$w_i = d_{i,\text{CENTER}}$	$w_i \leftarrow w_j$
VAM	$w_i = b_i - d_i$	$w_i = b_i - d_i$ (where d_i and b_i are now defined on the newly formed component)

5. ALGORITHM

Implementation

The above algorithm can be implemented in several distinct ways which vary significantly in their usage of computer time and memory. It is similar in structure to many published algorithms for computing MSTs and finding the shortest path between pairs of nodes in a graph. The techniques used are primarily those of Kershenbaum and Van Slyke [1972] and Johnson [1972], which take advantage of sparsity when it is present and are generally conservative of memory and running time. [See NAC Semiannual Report #5]

As we have already stated, the core requirements and running time can be greatly reduced by considering a solution within the framework of a sparse graph where each node is connected only to a few, say K, of its nearest neighbors and to the root. Even when K is relatively small, the solution thus obtained is not significantly worse than one obtained by consideration of a complete graph. Fig.4.3 shows the variation of line cost with number of neighbors considered in a 40 node network. As can be seen, there is virtually no increase in cost until the number of neighbors is reduced below 5 . This is not surprising in light of the fact that it is never advantageous to connect a node to another node which is further away than the root as the



TOTAL LINE COST VS. DENSITY OF SOLUTION GRAPH

FIGURE 4.3

connection to the root is always feasible and less expensive than a connection to the more remote node. Thus, such connections can be ignored without any danger of increasing the cost of the resulting network. For a network of nodes uniformly distributed over a region with the root near the center, roughly 75% of the possible connections can be eliminated in this manner. This, coupled with the fact that even with constraints, it is very likely that a node will be connected to one of its nearest neighbors in the optimal solution explains that the consideration of only a relatively small number of neighbors will yield a very good solution.

It is possible to find the K nearest neighbors of each node without even evaluating all internode distances. In a number of operations proportional to N, the number of nodes, the area containing the nodes can be partitioned into rectangles and the nodes in each rectangle identified. The K nearest neighbors of each node can then be found by considering the nodes in the node's rectangle and rings of the adjacent rectangle to whatever distance necessary. If the number of rectangles is chosen carefully, this can be done in a number of operations proportional to N x K, where K is the number of neighbors desired.

A potentially time consuming step in the algorithm is the recalculation of the t_{ij} every time an arc is brought into the

tree and the subsequent search for the minimum. We have already substantially reduced the amount of computation in this step by limiting ourselves to a solution within a sparse graph. Thus, instead of having to recalculate Nx(N-1)/2 values of t_{ij}, we need only recalculate NxK values.

The effort in recalculating the t_{ij}'s can be reduced still further by noting that they are in fact defined as a difference of two quantities, w_i and d_{ij}. The d_{ij} are constants and do not need to be recalculated and the w_i need, at worst, to be recalculated once for each node when an arc is brough in. Thus, the problem of recalculating NxK values of t_{ij} can be reduced to at worst recalculating N values of w_i if we are willing to have the t_{ij} represented implicitly by the values of d_{ij} and w_i. In practice, even this is usually not necessary. A w-rule such as that used by an algorithm like Esau-Williams' requires only the recalculation of the w_i for nodes in one component in the pair being merged. Prim's Algorithm requires only one w_i to be recalculated and Kruskal's Algorithm requires no recalculation at all.

If the neighbors of each node are kept in sorted lists and pointers maintained to the nearest neighbor of each node, the algorithm can be implemented with the t_{ij} represented implicitly

without any increase in computation beyond the additional subtraction of w_i from d_{ij}. Furthermore, if the current values of $\hat{t}_i = \min_j(t_{ij})$ are kept in a heap, the value of t^*_{ij} is always immediately obtainable from the top of the heap. The only computational expense incurred by such a procedure is the updating of the heap when an arc is brought into the tree. This is at worst a linear operation and is in fact proportional to $\text{Log}_2 N$ when the number of w_i charged at each step is small.

Thus, the entire algorithm is bounded by

$$\alpha N^2 + \beta KN + \gamma KN \text{ Log}_2 K$$

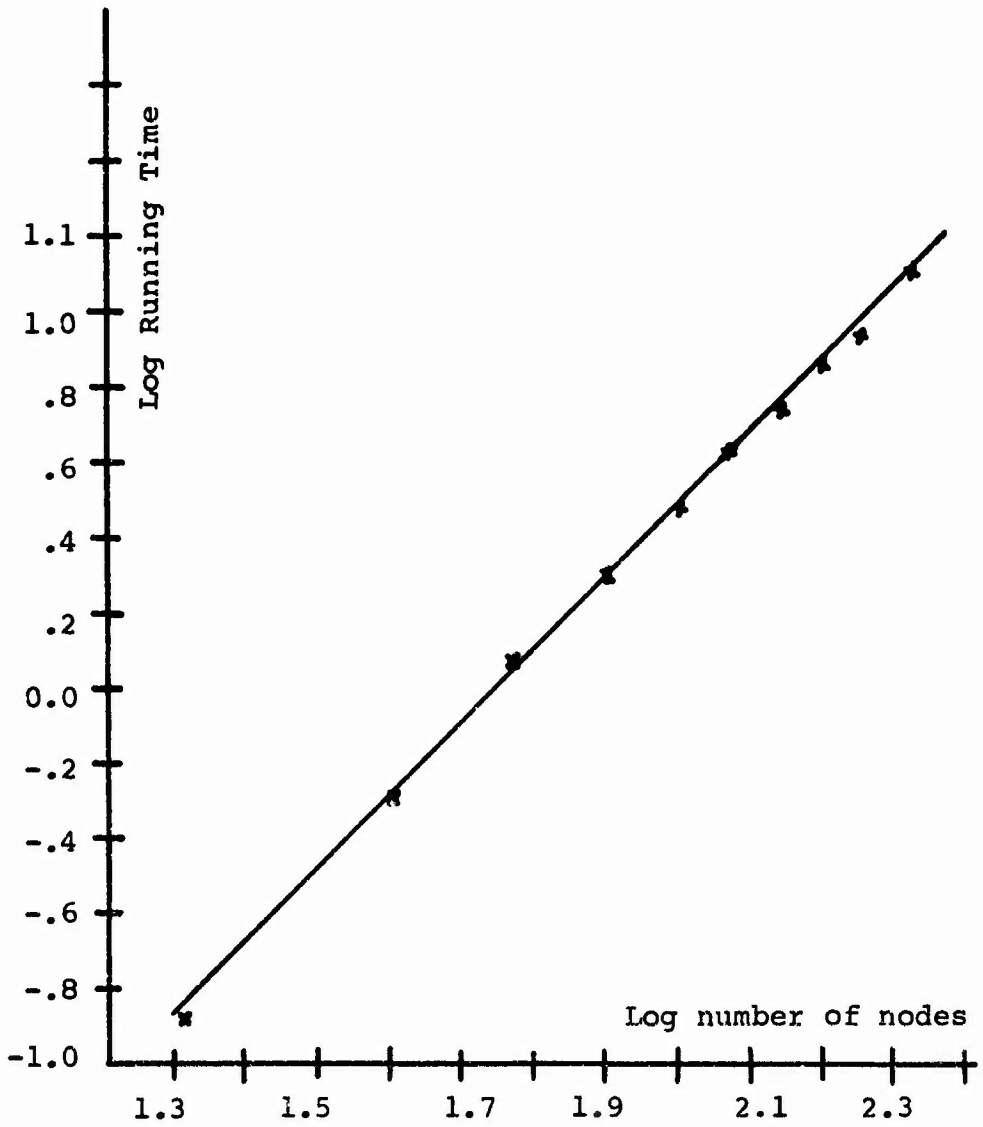
6. ALGORITHM PERFORMANCE

Experiments were performed using 20, 40, 60, 80 100, 120, 140, 160, 180, and 200 node networks considering 5 nearest neighbors, the w-rule for the Esau-Williams algorithm, and a traffic constraint of:

$$T \leq N/4$$

where T is the line traffic and N is the number of nodes in the network. Figure 4.4 summarizes the results of these experiments. The figure shows a plot of the logarithm of the number of nodes versus the logarithm of the running time. The data points fall almost perfectly onto a straight line with slope + 2, pointing out that the unified algorithm varies quadratically with problem size.

Other experiments verified that the running time did not vary noticeably with the particular w-rule used. Also, no appreciable variation in running time was observed when constraints were varied. Neither of these observations is surprising. Generally, in the implementation of a variety of w-rules, the program assumes that all the w's are reheaped each time an arc is brought in. This also accounts for the variation of running time with n^2. Constrained minimal spanning tree algorithms of this type examine roughly the same arcs during their execution,



RUNNING TIME VS. NUMBER OF NODES

FIGURE 4.4

stopping when they connect groups of nodes to the center. Thus, the running time remains nearly constant when the w-rules and constraints are varied.

Running time is, however, significantly altered by considering a smaller number of neighbors for each arc, and thus, a smaller number of arcs for possible inclusion in the tree. Table 4.2 shows running times for a 40 node network varying the number of neighbors considered. As can be seen, the running time can be reduced by a factor of 4 by considering 4 neighbors instead of 39, even in this relatively small 40 node case. For larger networks, the percentage of running time saved is even greater.

Comparison of Line Costs Using Known Heuristics

Experiments were run using Prim's algorithm, Kruskal's algorithm, Esau-Williams algorithm, and the VAM algorithm as specific subcases of the unified technique, i.e. input parameters were varied to implement the specific w-rules for these algorithms. Running times are not compared since within the context of the unified technique, they are nearly identical.

Twenty node networks with randomly generated traffic and coordinates were used in this comparison under a variety of constraints. Previous experience with these algorithms verified

TABLE 4.2

Number of Neighbors	Running Time	Line Cost
1	.434	7.77
2	.491	5.26
3	.522	4.54
4	.557	4.50
5	.589	4.48
8	.691	4.45
10	.767	4.45
15	.985	4.45
20	1.240	4.45
25	1.503	4.45
30	1.767	4.45
39	2.245	4.45

the fact that their relative performance is not affected by problem size. Twelve networks were run using each algorithm.

The results are summarized in Table 4.3. The Esau-Williams algorithm performs better than any of the others in all cases with the VAM algorithm producing results only slighly worse on the average.

These results should be interpreted in a qualitative sense rather than as an indication of actual percentage improvements obtained using one w-rule rather than another. Experience with a wide variety of problems has shown that the percentage variation between the solutions obtained using these w-rules varies greatly with problem size, constraints, and distribution of nodes. Indeed, in a few cases the VAM algorithm and even the others may actually yield a result superior to Esau-Williams. Therein lies the power of this unified approach. Not only can it easily handle a wide variety of constraints, but also, it can adapt itself to the particular problem at hand and yield results at least as good, and almost always better than any single known heuristic. Using it, one can apply several heuristics in succession or a hybrid technique using them simultaneously. The development of problem oriented w-rules based upon a generalization of those used in known algorithms is the subject of continued study.

TABLE 4.3

PERFORMANCE FOR 20 NODE NETWORKS WITH
TRAFFIC UNIFORMLY DISTRIBUTED BETWEEN 0 AND 1

	Case 1	Case 2	Case 3	Case 4	Average
Kruskal	4.04	5.09	3.49	4.33	4.24
Prim	4.41	5.13	3.62	4.77	4.48
Esau-Williams	3.65	5.06	3.28	3.97	3.99
VAM	3.71	5.15	3.43	4.14	4.10

Case 1: $N + T \leq 10$

Case 2: $N + T \leq 5$

Case 3: $T \leq 5$

Case 4: $N \leq 5$

N = number of nodes/line

T = total traffic/line

5. DISTRIBUTED ANALYSIS OF LARGE SCALE NETWORKS USING ARPANET

The development of a system of distributed computation in the design and analysis of large scale networks using ARPANET has begun at NAC. With the resources of TENEX systems at BBN and ISI, an IMLAC PDS-1D graphics minicomputer at NAC and recently, the addition of remote job entry use of UCLA-CCN's IBM 360/91, NAC is putting together a sophisticated software package of computational and interactive graphics programs for multicomputer editing, display and analysis of large scale networks.

The initial work toward this goal has been the development of an interactive graphics system, capable of displaying and editing networks. The project, written in a subset of FORTRAN, allows input both interactively from the user's terminal and from a file on a peripheral device for data and editing commands. The output is a file containing a NAC developed network data structure and sequences of graphics commands to the IMLAC for display. The resulting file can be saved for further editing, display, or computational analysis.

The procedure is as follows: Through the graphic display terminal (the IMLAC), the analyst constructs his network by

commands to a network editing routine. As the network is constructed, it is displayed dynamically, with appropriate node/link information for the analyst's reference. Upon completing the design and entering the required values for computations, the network is then analyzed by a specific program. On the basis of the results from that program, the analyst can modify the network, by removing and/or adding nodes and/or links, or input different values for computations and re-submit that new network for analysis. NAC's "Network Reliability Analyzer" is now on the ARPA Network and functions in this manner. To accomplish this:

1) Graphics routines had to be written to interface the PDP-10 and the IMLAC.

2) Input from the user's terminal had to bypass F40 (TENEX FORTRAN) input routines as it does not allow for unformatted I/O from that device.

3) A data structure had to be developed to conserve core, as a matrix network representation of large scale networks is wasteful of storage.

NAC received an excellent graphics driver package developed by Kevin R. Ray from the Computational Physics Group at the University of Utah. This, combined with NAC's network editing

algorithms, gives the user efficient and reliable graphics representation of the network design.

The formatted input is an inconvenience to the user, forcing the user to remember where his input is to go on an input record (whether in "columns" 1, 11, 21, or 1, 3, 9, etc.) To allow the user to determine "what" the input is to be, rather than "where" the input is to be, NAC developed a Free-Format routine to read from the user's teletype. This information is then converted into the desired internal format as requested by the program. With error correcting capabilities, this routine makes it easier, overall, for the user to enter the data required, and to answer the question posed by the program.

The network data structure is being developed dynamically as research into large scale networks continues. Functionally, it must be adaptable to both computation and display editing, as well as being compact for large networks. For this reason, the preliminary data structure has been designed in a matrix/list structure where fixed length properties of nodes and links are stored in the matrix, and variable length information (e.g., the list of links incident to a node) are in list structure. As further needs are realized, the data structure will be modified with a long term goal of designing the structure to represent

very large networks of general structure and characteristics. All network analysis and design programs for this study are compatible with this data structure and are being designed for using distributed computing.

Versions of each program will run from a user's teletype (i.e., no graphics support) or from the IMLAC (with the appropriate monitor) and have the capability of routing the computational work to the remote job service.

Immediate short term goals are (1) the implementation of NAC's network routing analysis program on ARPANET, using the newly developed interactive editing and display capabilities; and (2) the design of an appropriate programming language for the editing, display and analysis of very large network problems using a computer network.

6. PACKET RADIO

1. INTRODUCTION

A variety of studies have indicated that the key to extensive, successful use of computer communication and resource sharing networks will be access to flexible, efficient and economical local and regional communications. Present communication schemes are failing to meet these needs and consequently new approaches are necessary. One such approach, "packet radio", has recently been the focus of considerable effort by Network Analysis Corporation as part of its contractual responsibility to study cost-throughput-reliability tradeoffs in packet communication systems. In this chapter, we summarize some of the results of our initial studies. Details of these studies can be found in NAC's temporary working notes on Packet Radio listed in Table 6.1. As more is learned about the Packet Radio approach, the substantive portions of these notes will be rewritten and issued as permanent documents.

Component Tradeoffs

Stations in the Packet Radio System will be allocated on the basis of traffic. Thus, to first approximation we can think of partitioning the area to be covered into regions of equal traffic and allocate one station for each region. In regions of low traffic density the station may well not be in line of sight of

TABLE 6.1

Title	Date
Packet Radio Systems Considerations	1 Jan '73
Combinatorial Aspects of Flow in Packet Radio Nets--Part I	12 Jan '73
Comparison of Hop-by-Hop and End-to-End Acknowledgment Schemes	12 Jan '73
Packet Data Communications on CATV Systems	29 Jan '73
Channel Configuration for Packet Radio System	16 April '73
Data Options for Packet Communications on CATV Systems	5 April 73
Combinatorial Aspects of Flow in Packet Radio Nets--Part II	14 March '73

In Preparation

Time and Space Capture in Spread Spectrum

Channel Configuration--K-Station Model

Packet Radio Broadcast Network System Operation

Combinatorial Aspects of Flow in Packet Radio Nets--Part III

all the terminals in the region, hence repeaters are used to relay the traffic to the station. Thus repeaters correspond to a geographical partition of the area into sections small enough so that each terminal can communicate with a repeater and be relayed by it to a station. In areas of high traffic such as urban areas, repeaters will not be necessary; in fact the problem may be that a station can communicate with more terminals than it can handle.

2. SYSTEM CONSIDERATIONS

Structure

The Packet Radio System is a broadcast extension of a link * based packet communication system (such as the ARPA Computer Network) to accommodate various types of terminals without need of hardwire connections. The objective is to design an economic, reliable, and secure system for message communication in which Packet Radio Terminals communicate with Packet Radio Terminals as well as information processors on the link based network. The Packet Radio System will operate in a broadcast mode using the ALOHA random access method.

There are three basic function components: the Packet Radio Terminal (T), the Packet Radio Repeater (R), and the Packet Radio Station (S). Packet Radio Terminals will be of various types including TTY like devices, unattended sensors, small computers, display printers and position location devices.

The Packet Radio Station is the interface component between the broadcast system and the link based network. It will have broadcasting channels into the PRS and link channels into the link network. In addition, it will perform accounting, buffering, and directory and routing functions for the overall system.

The basic function of the Packet Radio Repeater is to extend the effective range of the terminals and the stations

*Link or link channel refer to point-to-point channels as opposed to broadcast channels.

especially in remote areas of low traffic and thereby increase the average ratio of terminals to stations.

The proposed system will therefore be composed of a terrestrial link communication network of the ARPA type, a terrestrial broadcasting network, and satellite channels in a link or broadcasting mode (see Figure 6.1).

The broadcast system approach is suitable for terminals which:

(i) are mobile so that a broadcasting mode is necessary,

(ii) may be located in remote or hostile locations where hardwire connections are infeasible,

(iii) have a high ratio of peak bandwidth to average bandwidth requirements, so that one uses to advantage the dynamic allocation of channel capacity without centralized control, or

(iv) require little communication bandwidth so that hardwire connections are uneconomical.

For large terminal densities, the stations will replace repeaters in providing area coverage to some extent; the extent to which this occurs depends on the distribution of population density. On the other hand, at high traffic levels repeaters play a much smaller role since the system is now traffic limited. In extreme cases repeaters may not even be necessary; however, there will be a very large number of stations and it will most

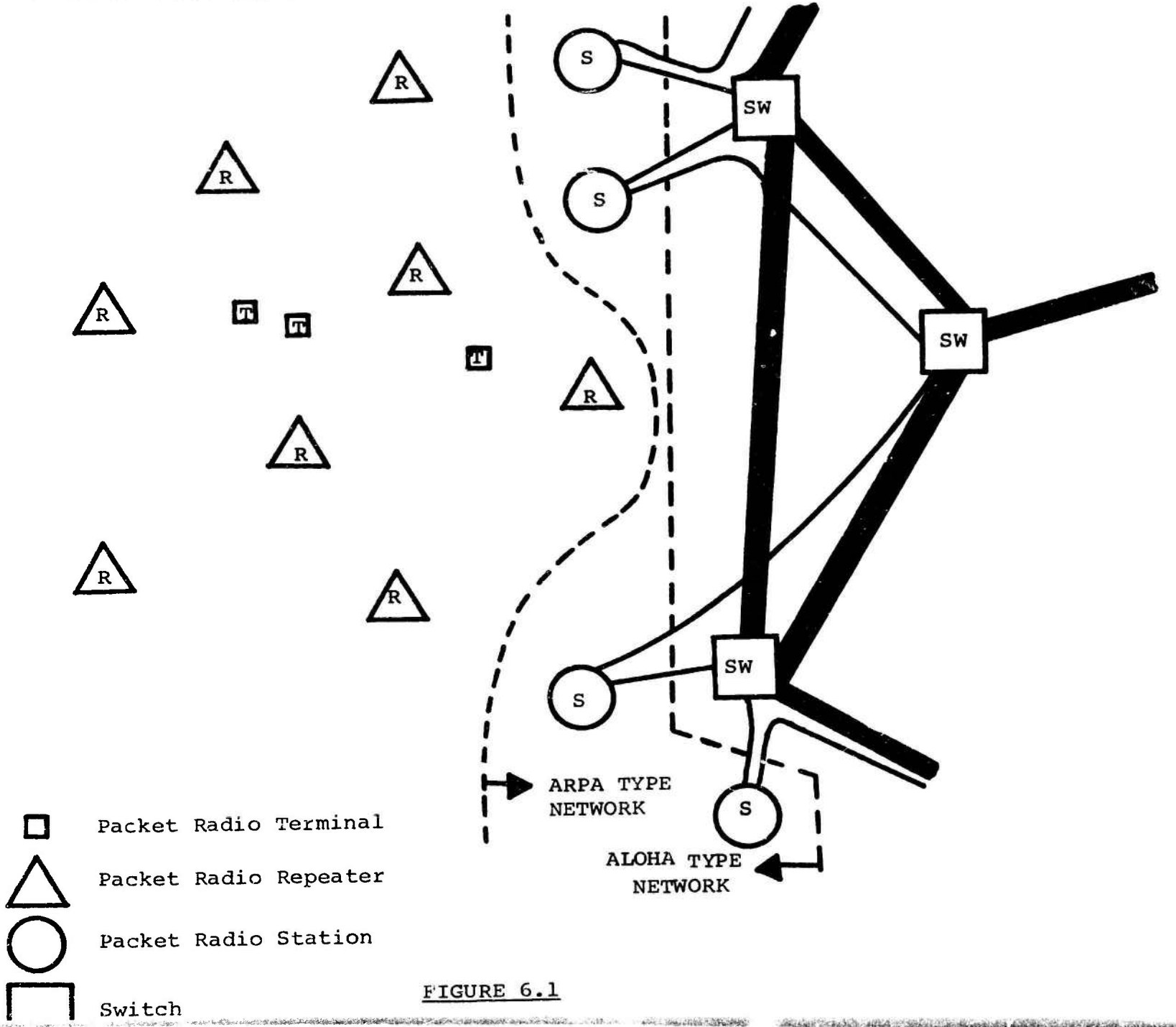

R
R
R
R
R
R
R
T
T
T
S
S
S
S
SW
SW
SW
ARPA TYPE
NETWORK
ALOHA TYPE
NETWORK
Packet Radio Terminal
Packet Radio Repeater
Packet Radio Station
Switch

FIGURE 6.1

likely be necessary to do some multiplexing or concentration before entering anything like the ARPANET.

A crude analysis leads to several preliminary conclusions. For low traffic levels, the number of stations << number of repeaters << number of terminals; hence, the assignment of functions to components should be such that the terminal is as simple and cheap as possible, the repeater only slightly more sophisticated and as many functions as possible should be delegated to the relatively few stations and the link packet communication network connected to them.

Many more factors affect the location of repeaters and stations than the simple ones indicated above. Terminal to station fanouts and repeater to station fanouts are affected by a variety of considerations. Moreover, considering repeaters simply as area covers and the station as traffic covers neglects important interactions between the two types of devices. Factors affecting the location of repeaters and stations besides range and traffic are:

(1) Logistics: certain locations for repeaters may be much preferrable to others since they may be more accessible or there may be available power so that batteries need not be used, e.g. on telephone poles or near power lines. Stations should preferably be placed near existing facilities of the associated ARPANET.

(ii) Reliability and redundancy: for a multitude of reasons redundant repeaters and stations will be required. Repeaters in remote areas will (very possibly) be operating on batteries which will fail and it would be necessary to provide sufficient redundancy so they will not need to be replaced immediately. Stations and repeaters will have intermittent and terminal failures for which backup is required. Extra repeaters are desirable when line of sight to the primary repeater is locally blocked. There will be random variations in repeater and station manufacture and placement which will cause inadequate or misdirected performance. This will have to be provided for by a safety margin of redundant coverage in the design.

(iii) Delays and throughput reductions due to collision and retransmission of packets: when a single channel is being operated in an ALOHA random access mode, no more than 1/2e of the bandwidth can be used in the unslotted case and no more than 1/e in the slotted case because of retransmissions resulting from packet conflicts. Spread bandwidth coding techniques may improve this figure but there will still be considerable extra traffic generated due to the repeaters and conflicts due to adjacent stations. The delays caused by retransmissions completely dominate the delays due to (i) and (ii) above. Some sources of retransmissions are:

(a) In order for the Packet Radio System to be reliable, it will be necessary, in general, that several repeaters or stations be within range of each terminal. If the repeaters retransmit everything they hear, one message can generate an exponentially growing number of relayed messages. In order to keep one message from saturating the whole network, some means of traffic control is required. The discipline chosen and its efficiency will probably be the single most important system factor affecting system performance. Two types of undesirable routing through the repeaters can occur. A message can circulate endlessly among the same group of repeaters if not controlled but even if this does not occur a message can be propagated in a geometrically increasing number of new repeaters.

(b) Again, for system reliability more than one station must be able to transmit via repeaters to each terminal. This means there will necessarily be conflicts between adjacent stations reducing the bandwidths from their nominal value and also introducing coordination and routing problems in the process.

(c) Because in general there will be many routes between any given terminal and any given station, many more conflicts will result than would be the case if the terminals communicated directly with a station.

Extraneous traffic can also be generated in the ARPANET if several copies of the same message enter the network from different stations. Either this can be sorted out on entry to the stations or at the destination. In the latter case, the traffic is artificially increased and in the former much more computation has to be performed by the stations to maintain coordination.

Component Functions

We discuss only the functional capabilities of the devices necessary for communicating in the Packet Radio network.

Terminals: There are two categories of terminals; (i) those which usually await a response to a message that they transmit (e.g. manually held radio terminals, small computers), and (ii) those which do not need any response or acknowledgment (e.g. unattended sensors, position indicators). Some terminals in the former category will usually send and/or receive several packets in one message.

The necessary capabilities of terminals in category (i) are:

(1) To identify whether the packet is addressed to their ID.

(2) To check whether the packet has a correct sumcheck.

(3) Some of these terminals will have character generation logic.

(4) Some of these terminals will have a random number generator when using random waiting time for retransmission; others may be assigned a random number for this purpose.

(5) Capabilities related to packet routing such as: terminating retransmission when acknowledged, recording and using a specific ID of a repeater and/or station to be used for other packets of the same message, counting the number of retransmissions.

(6) Capabilities related to the response to various, previously determined, types of error.

The capabilities of terminals of category (ii) are:

(1) Since these terminals are not operated by man they may have some functional capabilities by which a centralized control or a station will be able to identify whether the terminal is operative or dead. This may depend on the frequency at which the terminal transmits and the type of information.

(2) Those terminals which transmit "important" information or in general when it is important to receive all packets transmitted, should have the capabilities related to retransmission of the packet until acknowledged (see (4) and (5) above.)

Repeaters: The functional capabilities that repeaters should have are:

(1) To check whether a packet has a correct sumcheck and retransmit it.

(2) Capabilities by which a station can determine whether a particular repeater (or any repeater in a particular area) is operative or dead.

(3) When more "sophisticated" routing is used then repeaters should have the capabilities (1), (4), and (5) of terminals from category (i).

(4) Again, depending on the routing the repeaters may have additional capabilities related to determining the next repeater on the shortest path, capabilities related to labeling and re-labeling, etc.

Stations: The station will have a broadcast channel for communication with terminals and link channels connecting it to a nodal switch (SW) in the ARPANET. The switching machine may be similar in function to an IMP or a TIP. Every station will home on one SW with possibly a second for an alternative when the prime channel is down. It may be feasible that alternative channels will be used simultaneously under certain over load conditions.

(1) Cryptographic apparatus suitable for handling sensitive and private messages.

(2) A directory of terminals (and possibly repeaters) in its region.

(3) Operations necessary to convert short packets from ALOHA type network into long packets used in the ARPA type network.

(4) Storage buffers for packets received from terminals and packets to be transmitted to terminals.

(5) Storage for character position information for active terminals.

(6) Character generation logic.

(7) If station will be used to "connect" terminals in its region without going through the switched network, then it should have accounting capabilities.

(8) Capabilities related to routing of packets such as items (3) and (4) above.

Some of the above functions are optional and can be performed in the switched network (For example (2), (3), and (7)).

3. COMBINATORIAL ASPECTS OF FLOW

An immediate problem that arises in the successful operation of a packet radio system is the routing and control of packets within the network to achieve reliable and efficient operation. In order to study this problem, we construct an idealized combinatorial flow model on which the effect of different routing and control strategies can be tested. The models developed are described in detail in the documents "Combinatorial Aspects of Flow in Packet Radio Nets--Parts I and II" and in the forthcoming "Part III". In this section we summarize the problems studied in these documents.

It is assumed that repeaters are located at the corner points of a square grid depicted as follows:

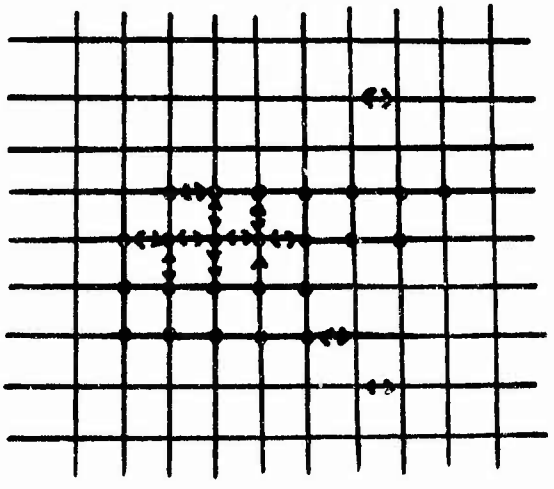


FIGURE 6.2

Under the first set of assumptions studied, a message received at any repeater is communicated (perfectly) to each of its four nearest neighbors, and received by those neighbors at the next time point.

Time is quantized in unit intervals (say one second). A packet that arrives within one interval is retransmitted within this interval. We initially omit consideration of packet length, channel utilization, and other propagation properties and effects. Our goal is to study message flow assuming all electronics have been properly designed.

We first analyze the effects of a single message originating at any repeater at time t=0. No other messages are introduced. The explosive effects of the single message are studied.

We then assume that messages arrive independently at each node at each point in time according to a Poisson distribution. That is, the probability that exactly j messages arrive at any node at each point in time during a unit time interval is given by:

$$P \text{ (exactly } j \text{ messages)} = \frac{e^{-\lambda} \lambda^{j}}{j!} \qquad j = 0, 1, 2, \ldots,$$

The parameter λ has the interpretation of the mean or average number of messages which arrive during a unit interval of time. In this case we determine the average number of messages received at a given point in the grid which for simplicity we call the origin (from Cartesian coordinates terminology).

We then study the probabilistic model introduced above under various possible operating modes of the repeaters. A measure of efficiency of the modes is introduced and calculated for the four modes considered. Each of the modes produce a reduction in redundant message flow. We next mix the operating modes and analyze the message flow. Analysis of the same models is also considered for the case where the grid is bounded and closed. The effect of a single message is determined by an algorithm for each point in time.

The results of the initial analyses lead to a concept of "inward labeling" in a finite grid structure with a station centered at the "origin". It is seen that by restructuring the size of the grid, substantial reductions in the numbers of copies generated by a single message are possible.

We consider in a detailed way the problem of message distribution when all messages are transmitted in the direction of the origin. The model now introduces the problem of conflict resolution and allows that some messages which arrive during overlapping time intervals at a repeater may not be repeated. When the message is not repeated we say it was not received. Thereby, we draw the distinction between arrivals and receptions.

Pictorially, the model can be described as a set of nodes or vertices which represent repeaters. The repeaters are at the integral lattice points of the plane, and the origin is considered the fixed station. Messages are repeated only in the direction of the origin. We can examine only the first quadrant due to symmetry. The arrows represent possible directions for a message.

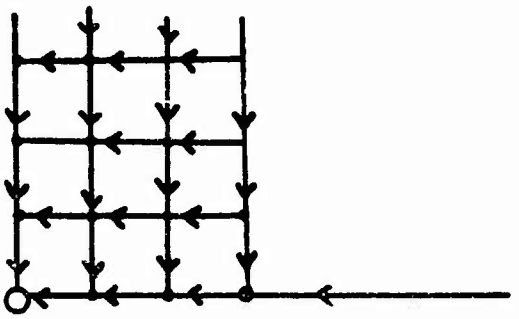


FIGURE 6.3

A repeater at distance d transmits to those repeaters one unit away which are at distance d-1 from the origin.

Basic Assumptions

(A) Starting at time t=0, and at quantized time periods afterwards (perhaps 1 sec.) t = 1, 2, ..., messages originate at each repeater independently according to a Poisson probability law. That is, the probability that exactly k messages originate is $\frac{e^{-\lambda} \lambda^k}{k!}$ when k = 0, 1, 2, ..., where λ is a constant and is the mean or average number of originations per unit time.

(B) Messages which are "received" at any repeaters are instantly repeated to all repeaters one unit closer to the fixed station. They arrive one time unit later at the neighboring repeaters.

(C) A repeater has the capacity to "receive" at most m-messages. As each message arrives at a repeater, it is randomly and independently assigned to a "slot"; there are m-slots. Two possible modes of operation are considered. In the first mode a message is "received" if it is the only message in its slot. In the second mode the number of messages "received" is the same as the number of non-empty slots.

A variety of specific questions are of interest in this model.

Problem 1

Find P_{kj} (j = 0, 1, 2, ..., k), k = 0, 1, 2, ..., which is the probability that exactly j messages are received given that k

arrive at a given repeater when type I "slotting" is used. That is, a message is received if and only if it is the only message assigned to its slot. In type I slotting, the number of messages received is given by the number of slots which have exactly one message. Note that P_{kj} is independent of time. Given that k messages arrive, the distribution of the number received does not depend on time or what happens at neighboring repeaters.

Problem 2

Find P^*_{kj} (j = 0, 1, 2, ..., k), k = 0, 1, 2, ..., which is the probability that exactly j messages are received given that k arrive at a given repeater, when type II "slotting" is used. In type II slotting, the number of messages received is given by the number of non-empty slots. The same remarks concerning independence hold for this problem as discussed above for Problem 1.

Problem 3

Find the probability $P_d(t)$ that a message which originates at time t, at distance d from the origin, is received at the origin at time t+d. The probability $P_d(t)$ is called the survival probability, and will be computed under type I, and type II slotting.

Problem 4

Let $X_o^R(t)$ be the number of messages received at the origin at time t, and $X_o(t)$ be the number of messages which arrive at the origin at time t.

Let $X_{j,d}^R(t)$ be the number of messages which are "received" at repeater with coordinates (j,d)* at time t, and $X_{j,d}(t)$ be the number which arrive, j = 0, 1, 2, ..., d, d= 1, 2, ...

All X's are random variables; compute their expectations and or distributions under type I and type II slotting.

Problem 5

Solve problems 3 and 4 when the region containing the repeaters is closed at say distance B from the origin.

Areas for Future Analysis

The results presented in the reports are only a beginning to the type of results that can be obtained using the same basic approach. There is the obvious broad general area of studying message flow at the origin as a function of the mean arrival rate ( ) of messages at any repeater, the number of slots per second (m) and the capture mode for the present model and its extensions. More specifically, we have not yet studied the probability of message survival as a function of distance of origination, time

*d is the distance (number of units of time) from the station, and j is the horizontal number of units.

of origination, λ , m and mode. This area is prime for our next analysis.

Another area yet unexplored with the present model is the effect of various types of breakdowns of parts of the network. This can be studied easily since all nodes are labeled and hence can easily be removed from the network calculations for an arbitrary desired length of time. The effect of such breakdowns can be analyzed in terms of all factors mentioned above as well as recovery time of the system.

The model can also be generalized and extended. Two specific directions are of interest.

(A) Message Flow from the Fixed Ground Station To The Repeater Network. We can repeat an analysis of the network when messages are flowing out from the origin. Questions similar to those posed for the inward model can be posed and analyzed.

(b) Message Flow In Both Directions. We can study the same model when messages are passing through repeaters to and from the origin. This network can be studied under various operating conditions which include one and two frequencies.

These studies are essential since they can reveal important properties of common channel and separate channel repeater operation as well as measures of reliability for proposed repeater nets.

4. ROUTING AND FLOW CONTROL

There are many possible paths a packet originating at a terminal may follow until it is received by the station. That is, a packet transmitted from a terminal can be received by several repeaters and that there may be several stages of repeaters before the packet is received by a station. Among the problems which arise in controlling traffic flow in a large scale broadcast network which one does not encounter in link communication networks are:

1) A packet transmitted can be received by many repeaters or stations or not be received by any.

2) Many copies of the same packet can circulate in the network.

3) Copies of the same packet can enter the ARPANET at different stations.

4) Different parts of one message can enter the ARPANET at different stations.

Some indication of the difficulty of the problem when no controls are imposed can be learned from one of the ideal cases discussed in the last section. In this ideal model the repeaters are located at corner points of a square grid at distances of one unit (time), a packet transmitted by a repeater is correctly

received (only) by its four nearest neighbors, a packet received by a repeater is immediately repeated. Suppose now that a single packet originates at the origin and that the transmission plus the propagation time is one unit of time, then after n units of time we have: (i) the number of repeaters which receive the packet for the first time, B(n), is:

$$B(n) = 4n, \quad n > 1; \quad B(0) = 1$$

(ii) the number of repeaters that the packet went through, A(n), is:

$$A(n) = \sum_{j=0}^{n} B(j) = 2n^2 + 2n + 1, \quad n \geq 0$$

(iii) if we assume that a repeater can receive and relay a large number of packets at the same time, then the number of copies of the same packet received by a repeater at coordinates (d,j) after d+2K units of time is:

$$N_j^d\,(d-2K) = \binom{d+2K}{k+j}\binom{d+2K}{k} \xrightarrow{\text{for large } K} 2^{4K}$$

where d is the number of units of time that the packet needs to arrive from the origin to the repeater, and j is the horizontal number of units. Thus, unless some steps are taken this explosive

proliferation of redundant packets will severely limit the capacity of the system. One can recognize two somewhat distinct routing and control problems:

(1) to assure that a packet originating from a terminal arrives at a station; preferably using the shortest path.

(2) to suppress copies of the same packet from being indefinitely repeated in the network either by being propagated in endless cycles of repeaters or by being propagated for a very long distance.

The following methods can be used for the suppression of indefinite packet propagation:

(A) A maximum handover number as in the hot-potato routing. Each time a packet is retransmitted the handover number in the packet is incremented by one. If the handover number exceeds an assigned maximum, the packet is dropped. If the maximum handover number is set to be large, extensive artificial traffic may be generated in populated areas; on the other hand, if it is set to be small, then packets from remote areas may never arrive at stations. This problem can be resolved as follows: We assume that every repeater knows its approximate distance to stations from observing response packets. The first repeater which receives the packet from a terminal sets the maximum handover

number by "knowing" the approximate radius in "Repeaters" in its region.

(B) Repeaters can save the header of packets (or possibly the entire packet) for a specified period of time to be compared with headers of packets (or with packets) received. If the same packet is received by a repeater the second time, it is not retransmitted.

In what follows we propose three routing techniques which can be used for broadcasting networks. In all methods it is assumed that a repeater knows whether a packet is addressed to a terminal or to a station. This is indicated either by the transmission frequency or by means of a bit in the header of the packet when the same frequency is used in both directions.

Method 1: Regionalization

The principle of this method is that every packet will contain several bits for a "regional address". This address is associated with one or more stations in a region, and possibly many repeaters.

Transmission from terminal to station: When a packet originates from a terminal it has a blank "regional address". If the packet is received by a repeater (with a correct checksum) and it has a blank regional address or the same regional address as the receiving repeater, it is retransmitted with the regional

address. If the packet received has a regional address which is different from that of the repeater, it is dropped. A station which receives the packet will transmit a display acknowledgment to the terminal. When a response packet is received by a repeater it is repeated, again with the regional address. The terminal will time-out and retransmit the packet if an acknowledgment has not been received with an indication that it is the same packet.

Transmission from station to terminal is the same as above.

When transmitting from a terminal to a station, it is possible that parts of the message will be received by several stations in the same region, or in different regions when the terminal is located at a boundary. To overcome this, it is possible to have several bits in the packet heater to indicate the station ID. After the terminal has received an acknowledgment for the first packet, it will transmit the other packets of the same message with the ID of the station. Other stations which received and acknowledged the first packet will save it for a "specified period of time" and drop it if more packets of the same message are not received.

Method 2: One Level Labeling

By this method one obtains shortest path routing in one direction, from the terminal to the station. In this method

and in Method 3, it is assumed that packets are routed using "hop-by-hop" transmission, i.e. a repeater stores the packet and keeps retransmitting it until acknowledged by the next repeater stage. It is also assumed that if device i can receive from device j, then device j can receive from device i.

For routing purposes every repeater i is characterized by the triple $(R_i, L_i, \overline{R}_i)$ where R_i is its identity number, L_i its label which indicates the number of hops on the shortest path to its nearest station, and $\overline{R}_i$ is the identity of the next repeater on the shortest path to the nearest station. For one level labeling R_i is fixed whereas L_i and $\overline{R}_i$ are modified to reflect changing conditions in the network.

The Packet Radio network is periodically relabeled by labeling packets from the stations, to adapt to a new state of the network. R_i relabels itself upon receiving a labeling number L_k from another repeater or station by:

$$L_i \text{ (new)} = \min[L_i \text{(old)}, L_k+1]$$

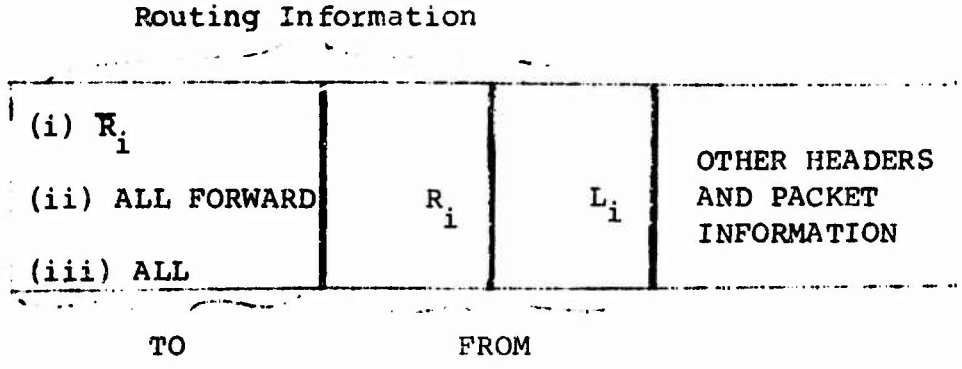
If L_i(new) = L_k + 1 then $\overline{R}_i$ is set to R_k.

When the network is labeled, then every repeater knows the next repeater on the nearest path to the station.

Transmission from Terminal to Station: The first packet (or a signalling packet) transmitted from a terminal is addressed to all repeaters. A repeater which correctly receives this packet

acknowledges it with its R_i and thus stops the retransmission of that packet. All future packets of the same message will be addressed to R_i and retransmitted by the terminal until acknowledged by R_i.

The packet transmitted from repeater to repeater up to the station includes the following routing information:



The packet is first transmitted to $\bar{R}_i$ for a specified number of times. R_i waits for a certain deterministic plus random time before each retransmission. If the packet is not acknowledged by R_i after the specified number of times then it is transmitted to ALL FORWARD (AF), again, for some (possibly different) specified number of times. AF means that the packet is addressed to all receivers with a smaller label. Thus when a repeater receives a packet with AF it has to check its label

against L_i. If the packet had not advanced with AF, then the next (last) step would be to transmit to ALL. ALL addresses the packet to all repeaters than can "receive" it. In particular it means that the packet can travel backwards and try a new path. When ALL is used it is possible, although with very low probability, that a packet will be transmitted in a cycle. Several procedures can be used for preventing the latter.

Transmission from Station to Terminal: The labeling of repeaters does not include sufficient inforamtion for directing transmission from the station to the terminal on the shortest path. This is particularly true when a packet is originally delivered to its destination.

To reduce the repetition of packets in this direction, however, one can use an address ALL BACKWARD (AB) by all repeaters. When AB is used by repeater R_i, the packet is addressed to all repeaters R_k for which $L_k > L_i$. This assures the suppression of packet repetition after it propagates once through the network. Another possibility is to regionalize the network.

A simplified version of Method 2 is without the labeling. The first packet originating from a terminal establishes a unique path through which other packets of the same message will

be transmitted. Packets can be addressed either to $\bar{R}_i$ or to ALL. The first packet from a terminal will be addressed to ALL. The repeater which received it will acknowledge it with its R_i. Other packets of the same message from the terminal will be addressed to R_i. R_i will respond only to packets addressed to it or to ALL; all other packets will be dropped. Repeaters are not mobile and can "learn" the location of the nearest station or repeater. Thus R_i can start transmitting to specific $\bar{R}_i$, and use ALL as the second option. A station will also respond only to packets addressed to it or to ALL. When a positive acknowledgment is received from more than one repeater or station then all, but the first, can be ignored. Note that this method does not guarantee the shortest path; however, it selects one which seems not congested at this point in time.

Method 3: Hierarchical Labeling

This method enables shortest path routing from terminal to station as well as from station to terminal. The packet header will contain sufficient information for determining the next repeater on the shortest path in each direction.

Consider the case in which the Packet Radio network is labeled as in Method 2. Then the network has an inherent hierarchical (tree) structure where every repeater "homes'

on the repeater or station from which it was labeled, as in the following figure:

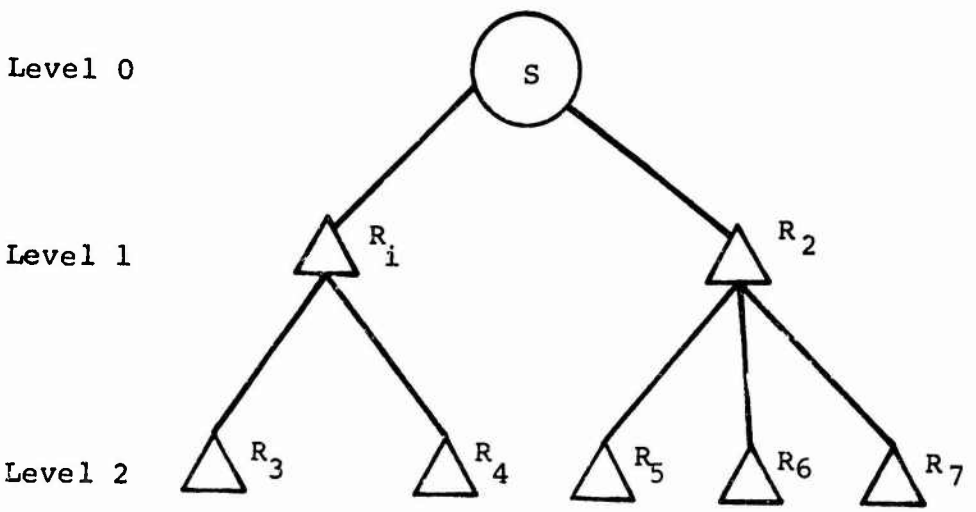


FIGURE 6.4

Suppose now that the identify number of each repeater is composed of h subfields, where h is the maximum number of hierarchy levels; and that repeaters which "home" on the same higher level (in the hierarchy) repeater are sequentially ordered. When this is done then shortest path routing can be obtained in both directions.

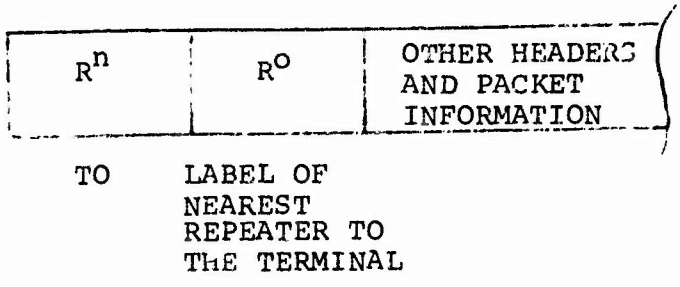
Every subfield has three possible entries--blank, a serial number, or ALL. The label of a repeater R_i in hierarchy level j will be composed of h subfields in which the last (h-j) subfields are blank.

As an example for the network in Figure 6.4, if we take 3 bits per subfield, then the identity numbers of the station and the repeaters can be as following:

	Subfield 1	Subfield 2	Subfield 3
S	0 0 1	0 0 0	0 0 0
R_1	0 0 1	0 0 1	0 0 0
R_2	0 0 1	0 1 0	0 0 0
R_3	0 0 1	0 0 1	0 0 1
R_4	0 0 1	0 0 1	0 1 0
R_5	0 0 1	0 1 0	0 0 1
R_6	0 0 1	0 1 0	0 1 0
R_7	0 0 1	0 1 0	0 1 1

In this example a subfield in which all bits are "0" is considered "blank". Note that all entries in Subfield 1 are the same since all repeaters home (eventually) on the same station.

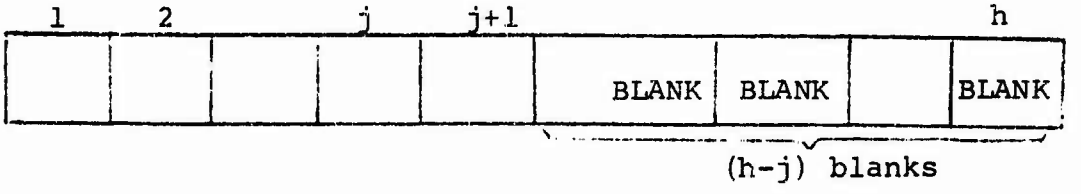
The packet header, in both directions, will include the following routing information.



R^o is the repeater or station which first acknowledged the first packet of the message, and R^n is the identity number, in this case also the address, towards which the packet is currently addressed.

Transmission from Terminal to Station: The first packet transmitted from a terminal is addressed to ALL. The repeater which acknowledges the packet also sends its identity number. This number is R^o. All other packets of the same message will be addressed by the terminal to R^o. R^o will be carried all the way to the station.

Suppose that R_i of hierarchy level j contains the packet. The identity number of R_i is as shown below:



R_i times out and retransmits the packet for a specified number of times to the repeater on which it homes. The address of this next repeater is the same as that of R_i, except that the (j+1)st subfield is set to be blank. If the packet is not acknowledged then it is addressed to AF, again for some specified number of times. There may be several stages of AF depending on how many

of the higher level subfields are set to "ALL". The next possibility is to address the packet to ALL. The address ALL is again similar to that of R_i where in the latter one replaces by "ALL" the subfields of j, j+1, and (j+2). The packet is then addressed to all repeaters whose hierarchy level is that of R_i, one level above or one level below. Again, there may be several levels of ALL depending on how many subfields above and below that of R_i are replaced by "ALL". Note that R_i does not need to know where the packet came from. Repeaters will respond only to packets addressed to their specific number, or if their specific subfield is ALL.

Transmission from Station to Terminal: R^o contains the information for the shortest path transmission to the terminal. The station transmits the packet to the repeater in which the first two subfields are specific (these of R^o) and all other blank. A repeater of hierarchy level j addresses the packet next, to the address in which the first (j+2) sublabels are from R^o and the remaining are blank. All other aspects of transmission, such as specific address, AB , or ALL, are similar to these in the reverse direction.

Note that the packet may depart from the shortest path when addressed to AB. When this happens then in the next stage the

packet tries to get back onto the shortest path by bypassing only the repeater that was busy or down.

To conclude, when the packet is routed from the terminal to the station and departs from the shortest path, it uses the shortest path from its new location; on the other hand, it tries to get back onto the original shortest path when routed to the terminal.

Remarks

(1) In Method 3, when the packet arrives to a station it contains the address of the "nearest" station to the terminal (the first subfield of R^{o}). It may be feasible to transmit this packet to its nearest station via link channels between the stations. If this is not done, then one may consider routing from terminal to station the same as from station to terminal; i.e. to try and get back onto the original shortest path.

(2) Many problems that may be encountered in routing and control of traffic flow are not discussed here. For example, algorithms for labeling and relabeling of repeaters are presently under extensive study and will be examined in a latter report.

(3) The "echo" type of acknowledgment where a terminal or repeater knows whether its transmission to a repeater is successful by listening to the latter repeaters transmit is not as effective

as in a satellite channel [See ASS Notes]. Because of FM capture the first repeater may be blocked from hearing the "echo" while the packet is successfully relayed. If a two frequency system is used, one for packets traveling towards the station and another for return, then the terminal will be listening to the return frequency while the echo is on the other frequency.

(4) If the ID to which a message is addressed cannot be found in its original location, then an information bank should be available where the new location can be found. The change of IDs' location can be recorded in a central or area-coded directly computer as proposed in [Roberts; 1972], or at the station of the original location as suggested in [Baran; 1964].

5. ACKNOWLEDGMENT SCHEMES

We consider the case where (n-1) repeaters separate the Packet Radio Terminal from the Station. Assuming that the terminal is at a distance of "one hop" from the first repeater one obtains the following n-hop system:

A simple model is used for evaluating the total average delay that a packet encounters in the n-hop system when using hop-by-hop and end-to-end acknowledgment schemes. When the end-to-end acknowledgment scheme is used, every repeater transmits the packet once. If the packet does not reach the station then retransmission starts from the terminal. The acknowledgment is sent from the station. In the hop-by-hop scheme repeaters store and retransmit the packet until positively acknowledged from the next repeater stage. Thus, one obtains "hop-by-hop" transmission.

The operation is so that a terminal, or a repeater in the "hop-by-hop" case, transmits the packet and if an acknowledgment does not arrive within a specified period of time, it retransmits the packet. The waiting period is composed of the time for the acknowledgment to arrive when no conflicts occur plus a random time for avoiding repeated conflicts.

Two different schemes for end-to-end acknowledgment and one scheme for hop-by-hop acknowledgment are studied. Curves for the total average delay as a function of the number of hops and the probability of successful transmission per hop are obtained. Two cases were considered: one in which the probability of success is constant along the path and another in which the probability of success decreases linearly as the packet approaches the station. Finally, channel utilizations are compared when using a slotted ALOHA random access mode of operation.

It is demonstrated that the hop-by-hop scheme is superior in terms of delay or channel utilization. This conclusion becomes significant when the number of hops increases or when the probability of successful transmission is low. For example, in a five hop system, if the probability of success per hop is 0.7 then the total average delay is 12.5 and 53 packet transmission times for the hop-by-hop and end-to-end acknowledgment schemes, respectively. The functional capabilities of the hardware required for using the schemes considered can be found in Section 2.

The model used is based on ASS Note 9 by L. G. Roberts and ASS Note 12 by L. Kleinrock and S. S. Lam. The model is simplified, however, by assuming that the probability that a packet is

blocked is the same when the packet is new or has been blocked any number of times before. Although the more general equations could have been written, the numerical solution is rather elaborate (see ASS Note 12) and seems unnecessary for this comparative study. It is further assumed that the probability of being blocked in different hops are mutually independent. By total delay is meant from the time that the first bit is transmitted by the terminal until the time that the last bit of the packet is correctly received by the station.

Notation

P = propagation delay per hop in one direction

T_f = transmission time of information packet

T_b = transmission time of acknowledgment packet

σ = rate of packets offered to receiver

γ = rate of packets with retransmission offered to receiver

q_i = P[successful transmission in hop i]

R = average waiting time before retransmission is made

$\overline{X}$ = average waiting time beyond the minimum to avoid repeated conflicts (we assume same for the different schemes considered)

D = E[total delay]

ρ = channel utilization

C = channel capacity

h, s = superscripts for denoting hop-by-hop and end-to-end transmission schemes respectively.

Above quantities with subscript i relate to i-th hop, unsubscript quantities refer to end-to-end or are the same for each hop, depending on the context.

$$\alpha = \frac{T_b}{T_f}$$

$$\beta = \frac{P}{T_f}$$

$$\delta = \frac{\bar{X}}{T_f}$$

The following schemes are considered:

D^h - hop-by-hop acknowledgment

D^{s1} - end-to-end acknowledgment where the waiting period before retransmission is composed of the time for the acknowledgment to arrive from the station plus some random time.

D^{s2} - end-to-end acknowledgment where the waiting time before retransmission is shorter; the same as in the hop-by-hop scheme.

We first examine the delays for the case in which the probability of successful transmission is the same for every hop along the path. The curves shown in the figures are for the parameters: α =0.5, β =0.02, and δ = 2.0. For example, β=0.02 occurs when d=15 miles for which the propagation time is 80 μsec, and when the packet transmission time is 4 msec (e.g. 400 bits @ 100 Kb/s).

Note that the curves show the delay as a function of q rather than channel utilization. Thus, they can be used for a slotted or non-slotted random access ALOHA Systems or possibly other modes of operation.

Figure 6.5 shows the normalized delay as a function of q, with n as a parameter. One can see that the delay for the end-to-end acknowledgment schemes grows much faster than in the hop-by-hop scheme. For example, in a 5-hop system, when T_f = 4 msec and q = 0.6 then the average delays are 68 msec, 188 msec, and 472 msec, for D^h, D^{s2}, and D^{s1}, respectively. Alternatively, for a 5-hop system and T_f = 4 msec, assume that an acceptable average delay is 40 msec (normalized delay of 10), then (from Figure 6.5) the lowest q which can be used are .92, .84, and .78 for D^{s1}, D^{s2}, and D^h, respectively. When a non-slotted random access ALOHA transmission system is used then the maximum effective utilization which can be obtain are 4%, 7.1% and 9.5%, respectively.

Figure 6.6 shows the normalized delay as a function of the number of hops n, with q as a parameter. Note that D^h is a linear function of n.

In practice q will differ along the path. It seems reasonable to assume that the probability of success, q, will decrease when the packet approaches the station. When a random access ALOHA

system is used, the practical range for q is from 1/e for which the effective utilization is maximum to 0.9 for which the utilization is 4.7% and 9.4% for the non-slotted and slotted case, respectively. We take a function of the form:

$$q_i = 0.9 - 0.5 \frac{i}{n} \; ; \quad i = 1,2, \ldots, n$$

The normalized average delay as a function of n with q_i a variable is shown in Figure 6.7.

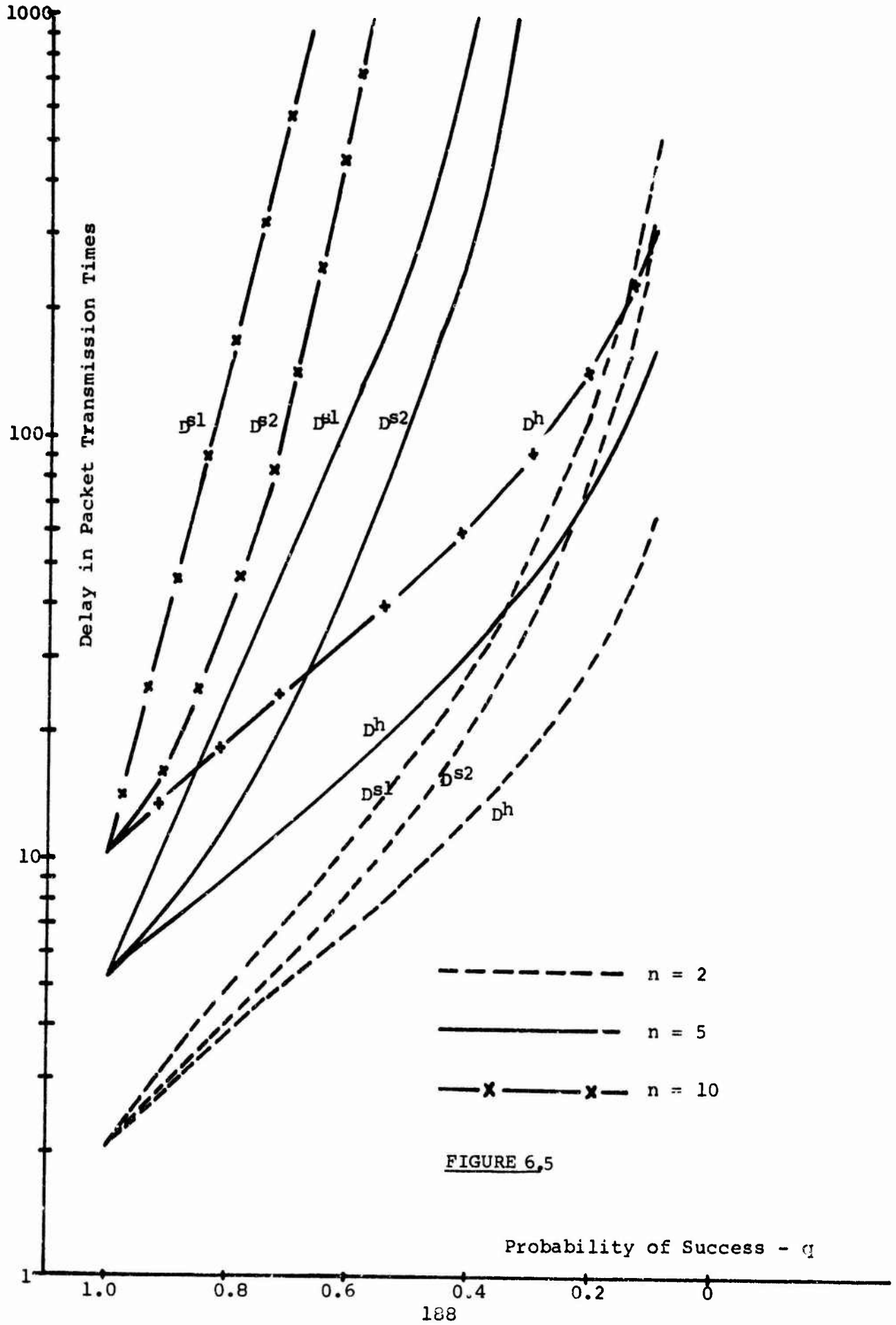

1000
100
10
1
Delay in Packet Transmission Times
D^{s1}
D^{s2}
D^{il}
D^{s2}
D^{h}
D^{h}
D^{s1}
D^{s2}
D^{h}
n = 2
n = 5
n = 10
1.0
0.8
0.6
0.4
0.2
0
Probability of Success - q

FIGURE 6.5

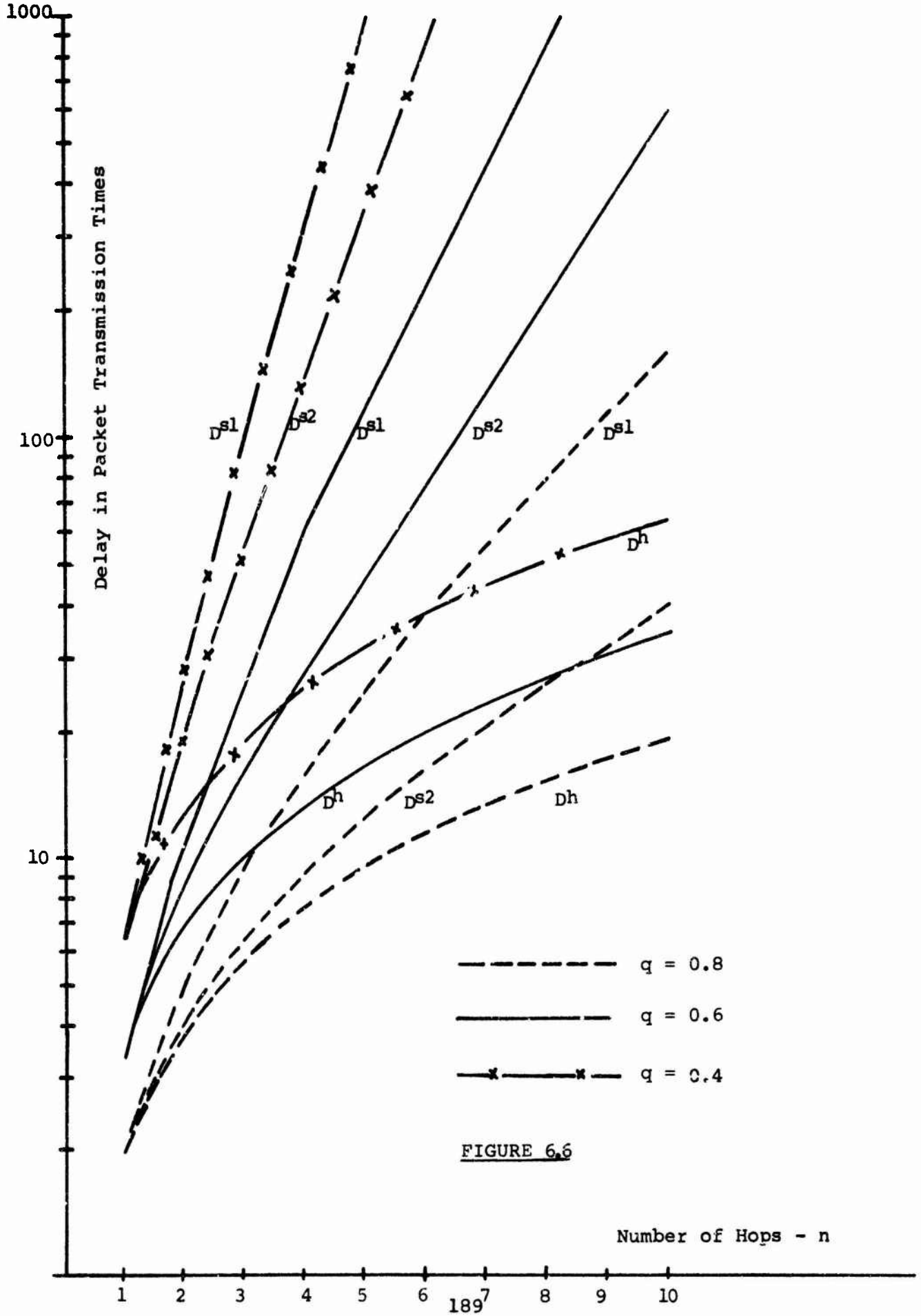

1000
100
10
Delay in Packet Transmission Times
D^{s1}
D^{s2}
D^{s1}
D^{s2}
D^{s1}
D^{h}
D^{h}
D^{s2}
Dh
q = 0.8
q = 0.6
q = 0.4
Number of Hops - n
1
2
3
4
5
6
7
8
9
10

FIGURE 6.6

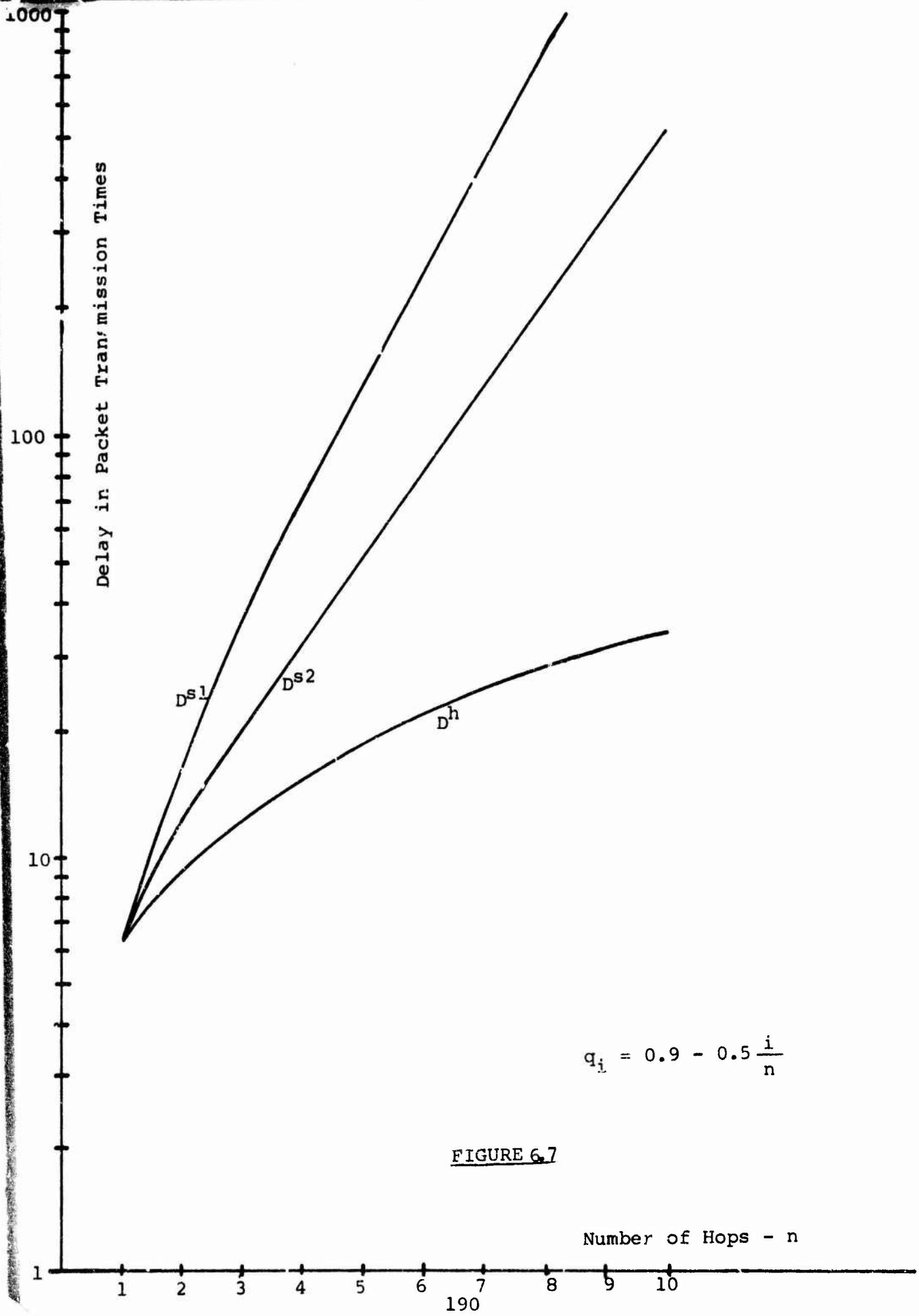

1000
100
10
1
Delay in Packet Transmission Times
D^{s1}
D^{s2}
D^{h}
$q_i = 0.9 - 0.5\frac{i}{n}$
Number of Hops - n
1
2
3
4
5
6
7
8
9
10

FIGURE 6.7

7. REFERENCES

1. Network Analysis Corporation Semiannual Report No. 3 for the Project "Research in Store and Forward Computer Networks," Defense Documentation Center, June 1971.

2. Network Analysis Corporation Semiannual Report No. 4 for the Project "Research in Store and Forward Computer Networks," Defense Documentation Center, June 1971.

3. Network Analysis Corporation Semiannual Report No. 5 for the Project "Research in Store and Forward Computer Networks," Defense Documentation Center, June 1972.

4. Network Analysis Corporation Final Technical Report for the Project "Research in Store and Forward Computer Networks," Defense Documentation Center, Dec. 1972.

5. P. Rosenstiel, "L'Arbre Minimum d'un Graphe," in Theory of Graphs, P. Rosenstiel, ed., Gorden and Breach, N.Y., 1967.

6. A. Kershenbaum and R. Van Slyke, "Computing Minimum Spanning Trees Efficiently, Proceedings of ACM Annual Conference, pp. 518-527, August 1972.

7. E. Johnson, "On Shortest Paths and Sorting," ibid, p. 510-517.

8. L. Roberts, "Extension of Packet Technology to a Hand Held Personal Terminal," Proceedings of SJCC, pp. 295-298, May 1972.

9. P. Baran, "On Distributed Communications," IEEE Transactions on Communication Systems, COM-12, pp. 1-9, 1964.

10. L. Roberts, ARPA Satellite Note No. 9.

11. L. Kleinrock and S. Lam, ARPA Satellite Note 12.